Multimodal estimation of discontinuous optical flow using Markov random fields

Fabrice Heitz, Patrick Bouthemy

UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1367

Programme 6
Calcul scientifique, Modélisation et
Logiciels numériques

MULTIMODAL ESTIMATION OF DISCONTINUOUS OPTICAL FLOW USING MARKOV RANDOM FIELDS

Fabrice HEITZ
Patrick BOUTHEMY

Janvier 1991

IRISA INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTEMES ALEATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CEDEX
FRANCE
Téléphone : 99.36.20.00
Télex : UNIRISA 950 473F
Télécopie : 99.38.38.32

Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields

Publication Interne n° 561 - Novembre 1990 - 50 Pages
Programme 6

Fabrice HEITZ and Patrick BOUTHEMY
IRISA/INRIA, Campus Universitaire de Beaulieu
35042 Rennes Cedex, France

Abstract

Visual motion analysis has become a subject of high concern in robot vision and dynamic scene analysis in the last decade. The estimation of dense velocity fields from image sequences is basically an ill-posed problem, primarily because the data only partially constrain the solution. Moreover, the optical flow estimation problem is rendered especially difficult by the possible discontinuities and occlusions between different moving objects in a dynamic scene. The classical regularization approaches to ill-posed problems usually assume smoothness constraints on the expected result. Because of occlusions, these constraints are not sufficient to yield satisfactory motion fields in the real-world case.

In this paper, we present a multimodal approach to the problem of motion estimation in which the computation of visual motion is based on *several complementary constraints*. The theoretical framework relies on global bayesian decision associated with statistical models, namely, Markov Random Fields. The multimodal algorithm addresses in parallel the estimation of velocity fields, the detection of motion edges, the processing of occlusions areas and the validation of several different motion-constraint equations. Local motion discontinuities are obtained as a by-product of the method, along with informations allowing to differenciate occluding regions from occluded regions. The estimation is gradient-based *and* feature-based and a local motion constraint takes part to the measurement only if it is acknowledged to be valid.

Results on synthetic as well as on real world image sequences are reported. Accurate motion measurements are obtained in the synthetic cases. The method compares favorably to other schemes on natural images including several moving objects with occlusions.

(Submitted to
IEEE Transactions on Pattern Analysis and Machine Intelligence
in November 1990)

Estimation coopérative du mouvement apparent par une approche markovienne

Résumé

L'analyse du mouvement apparent dans une séquence d'images est actuellement un des problèmes clefs de la vision dynamique par ordinateur. En particulier, le problème de la mesure du mouvement apparent apparaît comme une étape préliminaire indispensable dans nombre d'applications. Lorsqu'un champ dense de déplacements est souhaité, le problème est généralement mal posé (au sens d'Hadamard), principalement parce que les données n'imposent que des contraintes partielles sur la solution. L'estimation du mouvement est alors rendue particulièrement difficile par les discontinuités de mouvement qui apparaissent lorsque plusieurs objets aux mouvements différents évoluent dans la scène. Ces discontinuités s'accompagnent généralement de zones d'occlusion entre objets dans lesquelles toute information sur le mouvement est perdue. Les schémas classiques de régularisation des problèmes mal posés, qui imposent des contraintes globales de douceur sur la solution sont inadaptés dans ce cas.

En réponse à cette insuffisance, nous proposons ici une approche coopérative pour l'estimation du mouvement dans laquelle l'estimation d'un champ de déplacements dense repose sur un ensemble de contraintes différentes et complémentaires. Le cadre théorique de la méthode correspond à l'estimation bayésienne associée à des modèles statistiques globaux (champs de Markov). L'approche proposée permet, simultanément, d'estimer le champ des déplacements, de détecter les discontinuités de mouvement, de prendre en compte les occlusions et de sélectionner différentes mesures locales partielles du mouvement.

L'algorithme d'estimation du mouvement développé repose sur des techniques de relaxation. Il a été evalué sur de nombreuses séquences synthétiques et réelles. Les déplacements estimés se révèlent précis dans les cas synthétiques et se comparent très favorablement aux techniques classiques sur les images réelles.

1 Introduction

The recovery of visual motion from image sequences has been motivating number of investigations for the last decade, [2, 5, 31, 42] . The *optical flow field* can be defined as the distribution of apparent velocities of the brightness patterns in the image plane. As optical flow is related to the apparent spatiotemporal variations of the intensity function within the image sequence, it only imperfectly accounts for the real underlying velocity field due to the relative motion between the camera and the objects in the scene. This problem has been thoroughly addressed by Verri and Poggio, [53] , who have shown that the *optical flow field* is generally different from the *2D motion field* (projection on the image plane of the 3D velocity field of a moving scene). Nevertheless the discrepancies between the two fields are usually not so large and they both convey significant qualitative and quantitative information about the 3D environment, including relative depth, surface orientation, structure and motion of objects in space and sensor motion, [1, 2, 53, 54] . Dense optical flow computation thus appears as directly relevant to numerous problems in dynamic scene analysis such as moving object detection, [52] , motion-based segmentation, [1, 5, 41, 51] , qualitative kinematical labeling of moving objects in a scene, [17] , recovery of 3D motion and structure, [1, 8, 41, 54] with applications to robot navigation, obstacle avoidance, [17] or image coding, [31] .

Most velocity estimation methods belong to three main classes, [2, 5, 42] : gradient-based methods which derive optical flow from the spatiotemporal derivatives of the intensity function, [28, 29, 34, 43, 44, 46] , feature-based approaches which perform matching of image tokens over time, [6, 10, 13, 15] , and image transform techniques using spatiotemporal-frequency filters, [24] .

In this paper we are concerned with optical flow estimation methods providing *dense* motion measurements. The gradient-based methods lead to such dense fields whereas the estimates derived from feature-based schemes are generally sparse. It is well known that the estimation of dense velocity fields, like many other tasks in low-level vision, is an ill-posed problem. This means that the available data usually do not sufficiently constrain the solution of the problem. Additional smoothness constraints on the resulting motion fields are therefore to be introduced, [29, 45, 47] . Unfortunately, the classical answers to that problem suffer from several shortcomings.

As far as the motion constraints are concerned, the gradient-based measurements are known to be very sensitive to commonly encountered situations such as regions of constant intensity, motion discontinuities or occlusions areas. [34] . Large displacements are also beyond the scope of those methods. The need for other infor-

mation sources appears clearly here : one single motion constraint is not able to cope with the variety of real-world images. The smoothness constraints have also adverse effects on the estimated optical flow fields by blurring the motion discontinuities. Among those different problems the most difficult one is certainly the processing of the occlusions between different moving objects in a scene. The occlusions generate discontinuities in the optical flow field and give raise to regions in which no motion information is available. They may have adverse effects on the whole optical flow, due to the propagation of local information through the smoothness constraints. The usual computational approaches are not able to cope simultaneously with all these problems, because the constraints they introduce on the desired motion field only imperfectly account for the complexity of real-world scenes.

1.1 Multimodal motion estimation

In this paper, we present a multimodal approach to the problem of motion estimation.The computation of apparent velocity fields is based on several complementary constraints. The additional constraints introduced here aim to solve the different above-mentioned issues : optical flow estimation while preserving discontinuities, processing of occlusions and introduction of additional information sources. Local motion discontinuities are obtained as a by-product of the method, along with informations allowing to differenciate occluding regions from occluded regions. New constraints between motion discontinuities, intensity edges and the neighboring velocity fields are investigated. For the local motion measurement two information sources are considered : a gradient-based and a feature-based motion constraint. In our experiments, those two constraints revealed complementary. Gradient-based constraints [29, 34] provide motion measurements on dense sets and are well suited for textured images, with smooth spatiotemporal variations of the grey level distribution (natural outdoor scenes with small displacements often belong to this class). As aforementioned, they are sensitive to occlusion areas and large displacements and yield poor estimates on motion discontinuities and in highly textured regions [34, 53] . They are ill-conditioned in ambiguous regions of uniform intensity or in areas depicting linear varying intensity profiles in only one direction. On the other hand feature-based constraints [6] , based on edge-matching for instance, deliver only sparse velocity measurements but may give relevant results on motion and intensity discontinuities [6] . They are well adapted for indoor scenes or man-made environments exhibiting a reasonable number of prominent features. To combine those constraints properly, the validity of each constraint is locally tested and a given constraint takes part to the global estimation only if it has been acknowledged as valid. This avoids incorrect local constraints to contribute to the velocity field estimation.

Psychophysical investigations on visual perception have also shown that several

low-level information sources might be used by the human visual system, [28] . In [56] , Yuille *et al.* propose a visual perception theory, "the motion coherence theory" which leads to multimodal cooperation between different sources of motion measurement; yet, the occlusion problem and the computational aspects are not addressed.

1.2 Combining Information Using Markov Random Fields

In our approach, the theoretical and computational framework enabling a cooperation between several sources of information is based on the global bayesian decision theory and Markov Random Field (MRF) models. MRF models have been successfully introduced in several important low-level issues of static image processing such as image restoration, [4, 9, 20] , edge detection, [19] , image segmentation, [11, 19] , multisource image analysis, [27] , stereovision [3] , computed tomography, and surface reconstruction, [11, 38] . They have recently been extended to image sequence analysis, for motion detection, [7] , motion estimation, [25, 36] and motion-based segmentation, [41] . By defining a coherent mathematical framework for non-linear global statistical image modeling, they lead to significant improvement with respect to local methods. Markov Random Fields also appear as an efficient and powerful formalism for specifying non-linear spatial interactions between features of different nature, that is for combining information. By introducing strong generic knowledge on the features to be estimated in a given problem, they help to organize them, [19] .

In this paper the MRF modeling enables to jointly handle problems of optical flow estimation and issues of discontinuity and occlusion processing. The algorithm combines gradient-based and feature-based velocity measurements with evidences on occlusion areas in order to estimate dense velocity and discontinuity maps. Motion discontinuities are introduced in the velocity estimation and can be considered as a by-product of the estimation scheme. The occlusion regions (in which no local motion information is available) are explicitly taken into account in the global statistical model. Those regions benefit from motion information available in neighbouring regions, thanks to the MRF interaction model. Informations allowing to differenciate the occluding surfaces from the occluded one along a given motion boundary are also derived from the global markovian modeling. In regions where motion measurements are ill-conditioned, MRF appear as a natural way to regularize the estimation problem.

The resulting estimates correspond to the field which maximizes the distribution of the unknown velocity vectors and motion discontinuities, given the observed spatiotemporal variations of the intensity function. This relates to the Maximum A Posteriori (MAP) criterion in the framework of global bayesian decision.

1.3 Related Work

As far as motion estimation is concerned, Markov models were first used by Konrad and Dubois, [36] , to estimate smooth discrete valued velocity vectors via simulated annealing. In [32] Hutchinson *et al.* describe an analog and binary resistive network model equivalent to a Markov Random Field, to perform detection of motion edges simultaneously with the estimation of the velocity field. A VLSI implementation is derived. Konrad *et al.*, [35] and the authors, [25] have independently proposed to introduce binary edge sites between velocity vectors to estimate discontinuous motion fields. In [35] and [25] stochastic or deterministic relaxation algorithms are used to estimate the optimal velocity and edge configurations.

Preliminary results in a multimodal cooperation between different motion cues have been described by the authors in [26] . In a recent paper, Yuille *et al.*, [55] , develop a similar theoretical approach to integrate different disparity sources in a stereo algorithm, but no experimental results are reported.

Here, we propose new comprehensive MRF interaction models for optic flow estimation which differ from the work reported in [26, 32, 35] on the following points :

- The model can integrate different sources of motion measurements. It is illustrated here by a cooperation between gradient-based and edge-based motion measurements, but can be extended to a cooperation with other techniques (correlation, similarity functions, etc...).

- The model properly copes with the problem of discontinuity processing in image sequences. Following [26, 32, 36] motion discontinuities are modeled by *local* binary edges located midway between velocity vectors, but an additional feature of the model allows to take into account the *whole regions of discontinuity* corresponding to occlusion areas between objects undergoing different motion. This is a key-point since in real world sequences, taking into account false information within an occlusion region may lead to wrong velocity estimates and have adverse effect on the rest of the velocity field, [2] . It is demonstrated on several examples that the introduction of *local* binary motion edges, as previously proposed, is not sufficient to properly handle discontinuities in a moving scene.

1.4 Paper Overview

The multimodal motion estimation scheme described here is based on two complementary constraint equations : a classical gradient-based constraint [47] , leading to satisfactory estimates on dense points (apart from discontinuity locations), and a moving edge constraint, derived from a method presented in [6] , yielding sparse

velocity measurement on motion and sharp intensity discontinuities. The motion constraint equations are discussed in Section 2. For every motion constraint equation, we define a local validation factor for the corresponding measurement. Those validation factors show that low confidence is assigned to gradient-based measurements obtained near discontinuities (within occlusions, on sharp intensity edges and in highly textured regions) whereas the moving edge estimation method reveals reliable at most of those locations. The local motion information sources appear thus complementary.

Section 3 is concerned with the integration of the different constraints within a global bayesian decision framework. The unknown velocity field, and motion discontinuities are described through a coupled Markov Random Field model allowing smoothing of velocities, preservation of motion discontinuities, recovery of occluding and occluded surfaces, and cooperation between sparse and dense velocity measurements. The Maximum A Posteriori (MAP) criterion is adopted and leads to the maximization of the global a posteriori distribution of the motion features (velocity vectors and motion boundaries) given the observed grey level variations in the analyzed sequence. In the context of MRF modeling, the MAP criterion reduces to the minimization of a global energy function depending on the unknown velocity and discontinuity fields. An approximation to the MAP estimate is computed by a deterministic relaxation algorithm with good experimental results. Though this approach remains cpu time consuming on serial computers, the involved local computations are well adapted for an implementation on digital or analog massively parallel architectures, [32] .

Results on synthetic as well as on real world cases are presented in Section 4. The velocity field computed from synthetic sequences show significant improvement over the output of classical gradient-based schemes, especially in occluded areas. Accurate motion measurement as well as relevant motion discontinuities are extracted from real world sequences.

An extension of the multimodal scheme to multiscale motion estimation is described in Section 5. The multigrid algorithm enables the estimation of large displacements which are beyond the scope of the gradient-based techniques. The reduction in computation load is also appreciable. The multiresolution estimation method used here is coarse to fine and leads to an incremental estimation of the optical flow field. The multigrid estimation scheme performs well when large displacements exist in the moving scene or when it is necessary to propagate motion information through large uniform areas.

2 Multiple Motion Constraints

The multimodal motion estimation scheme relies on two motion measurement constraint equations that will be defined hereafter. This differs from the so-called "multi-constraint" methods which rather consider several inputs to the same equation, [40] , whereas our multimodal approach is based on a cooperation between different complementary motion constraints. This is a key-point, since problems often arise in the "multi-input" methods because the used inputs (multispectral data for instance) do not supply real complementary information, leading to ill-conditioned systems, [5] .

The first constraint under concern is the classical motion constraint equation proposed by Horn and Schunck [29] , and the second one is related to a moving edge estimation method recently described in [6] . The reliability of these two different motion constraints is discussed and validation factors are associated to both equations. In [34] , Kearney *et al.* have undertaken an error analysis of gradient-based estimation schemes involving local optimization. By surveying the different possible sources of error, they obtained accuracy and reliability measures on the estimates. We propose here an alternate approach that does not require a detailed study of every error source. It is based on hypothesis testing techniques and concerns both the gradient-based method and the feature-based one. The likelihood of the measures, assuming some underlying statistical models, are used to derive validation factors for the computed estimates. Those validation factors weight the contribution of the different constraints in the global interaction scheme described later in Section 3.

2.1 A Gradient-based motion constraint

Let $f(x,y,t)$ denote the observed intensity function, where (x,y) designate the 2D spatial image coordinates and t the time axis. A first order Taylor expansion of $f(x,y,t)$ leads to the following decomposition :

$$\begin{aligned} f(x+dx,y+dy,t+dt) &= f(x,y,t) + dx\, f_x(x,y,t) + dy\, f_y(x,y,t) \\ &+ dt\, f_t(x,y,t) + O^2 \end{aligned} \quad (1)$$

(where f_z denotes $\frac{\partial f}{\partial z}$).

The gradient-based motion constraint relies on two fundamental assumptions :

- the higher order terms in decomposition (1) are negligible ; that is, the intensity pattern can be locally approximated by a linear function of time and space ;
- the brightness of a point in space is supposed to be invariant between time t and $t+dt$.

Let $\vec{\omega}_s(u_s, v_s)$, $u_s = \frac{dx}{dt}(s), v_s = \frac{dy}{dt}(s)$ denote the velocity vector at point $s = (x, y, t)$; equation (1) and these two assumptions yield, [47] :

$$f_x(s).u_s + f_y(s).v_s = -f_t(s)$$

or, equivalently :

$$\vec{\nabla} f(s).\vec{\omega}_s + f_t(s) = 0 \tag{2}$$

where $\vec{\nabla} f$ is the spatial image gradient, $\vec{\nabla} f = (f_x, f_y)$.

Relation (2) is known as *the image flow constraint equation.* This analytical development outlines that the only velocity component parallel to the spatial image gradient can in general be recovered through local computation[1] . This is referred to as the *aperture problem.* In order to derive the complete velocity vector, it is usually assumed that points in the neighborhood of a given point move with similar velocity. *Local optimization approaches*, [34] , assume constant velocity in the neighborhood whereas *global optimization techniques* rely on a smoothness assumption of the velocity variations over the whole image, [29] .

The limitations of the early gradient-based techniques appear clearly and can be stated as follows :

- The surface shading may change over time (even for simple 3D movement such as rotation in space for instance). Obviously, this violates the constant brightness assumption.

- A first order approximation of the moving intensity pattern is an important source of errors, [6, 34, 53] . Indeed, such a linear approximation is relevant only as long as a relationship between the local spatial extent of intensity variations and the observed temporal changes is maintained by the displacement. This relationship no longer exists in occluded regions, or on motion discontinuities and also on intensity discontinuities (on sharp edges, or in highly textured regions for example). Large displacements are also beyond the scope of those methods for the same reason.

- The gradient-based schemes are sensitive to ambiguous areas such as uniform regions or regions exhibiting a linear variation of the intensity in one direction only.

- In real world images, the velocity fields are neither locally constant nor globally smooth : they are rather *piece-wise* continuous.

[1] This is not the case for grey level patterns such as grey value corners or grey value extrema (see [44]).

In practice, however, the existing schemes show (limited) robustness to those different sources of error, mainly because they minimize some error function with respect to the underlying imperfect model.

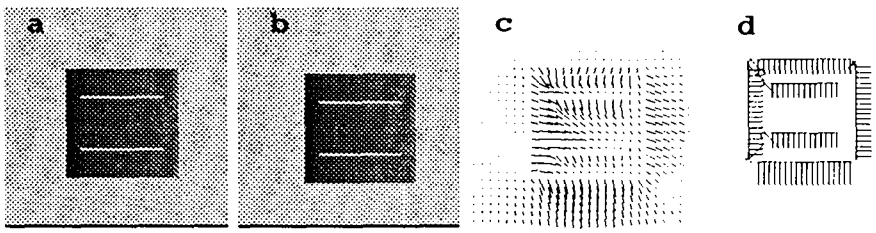


Figure 1 : motion estimation on the square sequence.
a)-b) Original sequence (100 x 100) : the square undergoes a translation of (2,2) pixels in the image plane. White noise has been added to the background.
c) Velocity estimation using a gradient-based method, [29] .
d) Velocity estimation using the moving edge estimator described in [6] .

Horn and Schunck gradient-based method assuming global smoothness of the velocity field, [29] , is illustrated on a synthetic image sequence (Fig. 1a-b). The sequence exhibits strong intensity discontinuities inside the square, along with occlusion areas corresponding to parts of the background covered or uncovered by the moving pattern. These occlusion areas create motion discontinuities on the square boundaries. As expected, the resulting velocity field (Fig. 1c) is blurred across the motion discontinuities. Moreover, poor quality estimates can be observed both on the central intensity lines and in the occlusion areas, due to the different limitations described above.

Different solutions have been suggested to cope with these problems. As far as variations in surface shading are concerned, in [12, 22] the gradient-based equation is generalized in the case of (limited) changing in the illumination conditions. The problem of discontinuities in the motion field is considered by Nagel and Enkelmann, [45] , who use an *oriented smoothness constraint* which prevents smoothing of the velocity field in directions where important variations of gray values are detected. Schunck, [48] , investigates clustering of local gradient-based constraints in order to obtain homogeneous motion measurements. The detection of motion discontinuities has been considered in several recent papers as a fundamental issue in motion estimation. Two approaches have been studied : the first one detects discontinuities *after* computing the optical flow, [51] , (for a review see [49]), the second one addresses the detection problem *prior or simultaneously* with the motion field estimation, [5, 25, 28, 32, 35, 37, 49] . Techniques of the second class give better results because the prior knowledge of motion boundaries helps to prevent velocity

smoothing through regions undergoing different movements. In [5] , Bouthemy *et al.* perform motion-based segmentation on partial motion information, using a likelihood test within a split and merge procedure. Spoerri *et al.*, [49] and Little *et al.*, [37] , propose several occlusion detection techniques based on the analysis of the behaviour of matching algorithms in the vicinity of motion boundaries. Hutchinson *et al.*, [32] and Konrad *et al.*, [35] consider binary motion edges in the estimation process and detect motion discontinuities *simultaneously* with the velocity field measurement.

In [25, 26] , the authors have independently introduced a similar modeling of motion edges with several significant extensions. A likelihood test has been developed to take into account occlusion regions in the estimation of the velocity field, [26] . The results obtained in [26] with this additional information, compares favorably with the other schemes, especially when occlusions are important. This approach will be generalized in Section 2.3 by testing the validity of the different motion constraint equations.

2.2 An edge-based motion constraint

We consider here a second complementary motion constraint which is *feature-based.* The underlying estimation method has been recently described in [6] . It will be called the "moving edge (ME) estimator" and is based on spatio-temporal surface modeling and hypothesis testing techniques. It simultaneously yields from some local processing the following output concerning moving intensity edges :

- edge position d
- edge orientation θ
- velocity vector perpendicular to the edge $\vec{\omega_d}^{\perp}$.

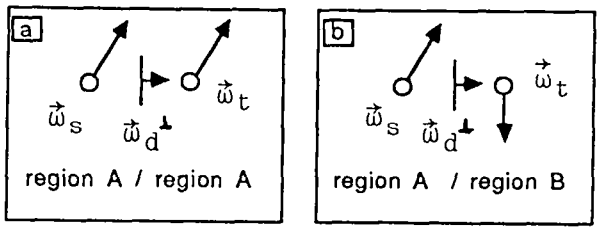
To this end, a spatiotemporal edge in an image sequence is modeled as a surface patch in a 3D spatiotemporal space. Within an elementary volume π in this 3D space (x, y, t), two local configurations may be encountered : either there is no spatiotemporal edge inside π or there is one. Two competing hypotheses H_0 and H_1 are associated to those configurations, [6] :

- H_0 : there is no ME ; then the intensity distribution within π is modeled as constant level c_0 + *noise*, where the noise is assumed to be a zero-mean Gaussian noise with variance σ^2.

- H_1 : there is one ME; then a surface patch denoted by $S(\Phi)$ subdivides π into two subvolumes π_1 and π_2 and the intensity distribution is modeled as previously, but according to constant level c_1 within π_1 and c_2 within π_2, with $c_1 \neq c_2$.

The log-likelihood ratio is optimized with respect to the pameters Φ of surface S, and the intensity parameters c_0, c_1 and c_2 under each hypothesis. Edge location, orientation and displacement are directly related to the optimal parameters $\hat{\Phi}$ of the determined surface patch if present. To decide whether a ME is present or not, the log-likelihood ratio is compared to a threshold λ (we refer the reader to [6] for more details).

The resulting test integrates both detection of moving edges and determination of their spatiotemporal characteristics. The local motion measurement is reliable, even on motion discontinuities and for important displacement magnitudes. Due to the aperture effect only the perpendicular component of the displacement can be derived.



a) moving edge within a region
b) occluding edge between two regions
| : moving edge site O : pixel site

Figure 2 : Location of moving edges with respect to pixels for a vertical edge. (the case of horizontal edges is similar).

In the ME estimator version we use, edge sites d can be considered as located midway between pixel sites (Fig. 2). To simplify the optimization of the likelihood ratio, edge location is given by a spatial intensity edge detector. This early step avoids maximizing with respect to edge location and orientation in the ME estimator, thus saving computation time. As a matter of fact, we use a recursive filter derived from Canny's criterion and designed by Deriche, [14] .

The velocity component perpendicular to the edge at location d constrains the unknown nearest neighbor vectors $\vec{\omega_s}$ and/or $\vec{\omega_t}$ through a projection equation called *the moving edge constraint equation* (the notations refer to Fig. 2) :

$$\vec{\omega_s} \cdot \frac{\vec{\omega_d}^{\perp}}{\| \vec{\omega_d}^{\perp} \|} - \| \vec{\omega_d}^{\perp} \| = 0 \qquad (3)$$

where $\vec{\omega_z}$ designates vectors $\vec{\omega_s}$ or $\vec{\omega_t}$, and $\frac{\vec{\omega_d}^\perp}{\|\vec{\omega_d}^\perp\|}$ is the unit vector normal to the intensity edge. The moving edge constraint states that the projection of the unknown velocity $\vec{\omega_z}$ on the unit vector perpendicular to the moving edge is equal to the norm of the perpendicular component $\vec{\omega_d}^\perp$.

If the detected moving intensity edge is related to an occlusion between two different regions, the constraint only holds for the velocity vector belonging to the same region as the occluding edge. In the case of Fig. 2b for instance, it turns out that the constraint only holds for vector $\vec{\omega_s}$, and not for vector $\vec{\omega_t}$. In order to propagate the constraint properly, it is necessary to determine to which region the occluding edge belongs. An information allowing to differenciate the occluding region from the occluded one is therefore required. This will be defined later in the global markovian modeling (Section 3.1). It will be considered as an additional, unknown feature in the model, and will be estimated in parallel with velocities and motion discontinuities.

Fig. 1d presents the result of the moving edge estimator on the *square sequence*. One can point out that unlike the gradient-based approach (Fig. 1c) the moving edge estimator yields good measurements on motion and intensity discontinuities (however motion information remains obviously sparse and only perpendicular velocity components are recovered).

2.3 Reliability of motion constraints

The accuracy and the reliability of the partial measurements associated to motion constraints (2) and (3) depend basically on the adequacy between the observed variations of the intensity pattern and the spatiotemporal changes the underlying model accounts for. For the gradient-based constraint the underlying model is essentially a linear model of the spatiotemporal pattern in space and time, with constant brightness assumption (see Section 2.1). The moving edge constraint relies on spatiotemporal surface modeling and parametrization.

For the different existing motion constraint equations, little attention has been paid to get reliability measurements (apart from [6, 34]). In the following we define validation factors associated to the constraints we have introduced. They will be used to withdraw the contribution of invalid local constraints from the global estimation.

2.3.1 A validation factor for the moving edge constraint

As far as the moving edge constraint is concerned, the ME estimator itself provides a natural way to define a validation factor. Let us recall that the determination of a moving edge at location d leads to compare the log-likelihood ratio associated to two competing hypotheses to a threshold (see Section 2.2 or [6]). The optimal value of the log-likelihood ratio L_d at location d, with respect to surface parameters

$\hat{\Phi}$ and intensity parameters $\hat{c}_0, \hat{c}_1, \hat{c}_2$ can be used to measure the reliability of the corresponding moving edge.

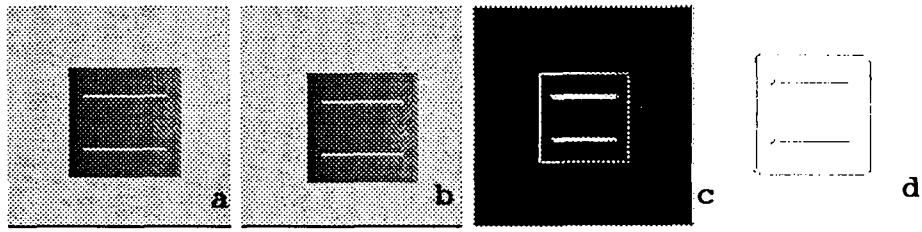


Figure 3 : confidence factor on the "square sequence".
a)-b) Original sequence (100 x 100) : the square undergoes a translation of (2,2) pixels in the image plane.
c) Log-likelihood surface for the ME estimator.
d) Binary confidence factor $\xi_{me}(d)$

Fig. 3c shows the log-likelihood surface in the case of the moving square. The likelihood surface indicates high reliability in the vicinity of edge location. We introduce following binary-valued validation factor ξ_{me} :

$$\begin{aligned} \xi_{me}(d) &= 1 \quad if\ L_d(\hat{\Phi}, \hat{c}_0, \hat{c}_1, \hat{c}_2) > \lambda_1 \\ \xi_{me}(d) &= 0 \quad else \end{aligned} \tag{4}$$

where λ_1 is a threshold.

The validation factor then gives the location of the most reliable moving edges (Fig. 3d).

2.3.2 A validation factor for the image flow constraint

Basically the image flow constraint (1) holds as long as the local intensity variations and the observed temporal changes are related. This property refers to the local spatiotemporal linearity and derivability of the intensity function.

Fig. 4a and 4b depict 1D intensity profiles $f(x,t)$ undergoing a displacement along the x axis between time t and $t+1$. In the case of Fig. 4a, the spatial variation corresponding to the displacement Dx is related to the temporal variation Df at point s through the image flow constraint :

$$\frac{\partial f}{\partial x}(s).Dx = -Df$$

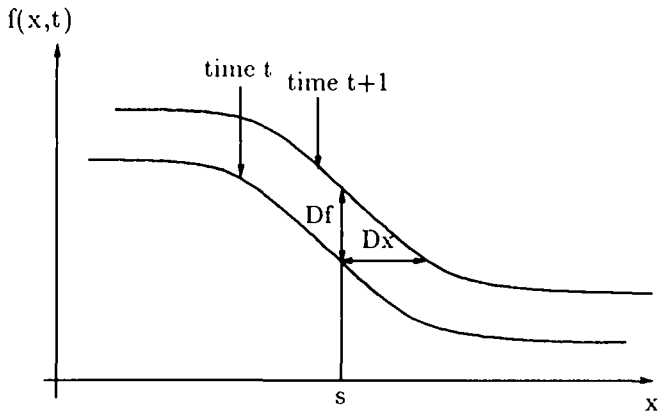


Fig. 4a

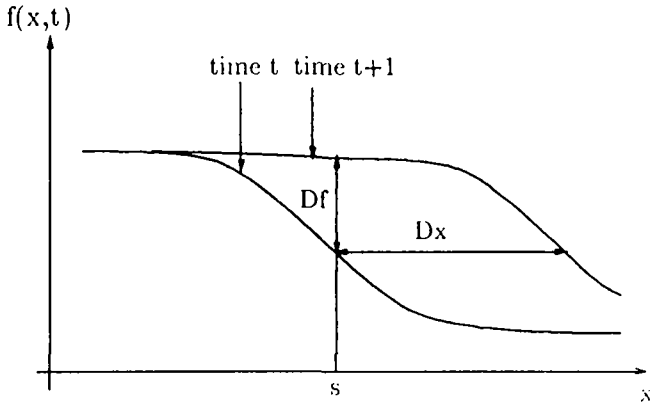


Fig. 4b

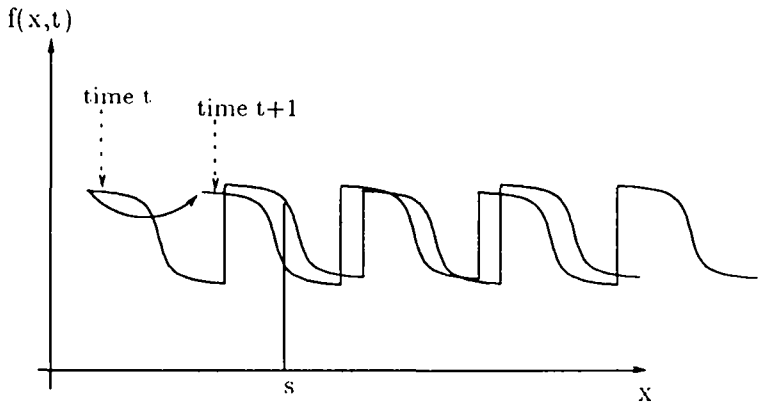


Fig. 4c

Figure 4 : Reliability of the image flow constraint.
a) Situation where the constraint holds.
b) Due to a large displacement, the constraint is broken
c) Ambiguous pseudo-periodical pattern

This is not the case in Fig. 4b where this relationship is no longer valid because the displacement is too large (a similar situation occurs when an occluding edge or a sharp intensity edge is present).

To determine whether the observed variations preserve the spatiotemporal relation, it is sufficient in practice to test if the first order spatial derivatives of the intensity function at point s remain the same between time t and $t+1$. This test does not account for situations such as the pseudoperiodical intensity pattern depicted in Fig. 4c, or similar ambiguous situations, which anyway are beyond the scope of most methods.

The hypothesis test considered here relies on a local linear model for intensity function $f(x,y,t)$ at point $s=(x,y)$ in two successive images (t and $t+1$) of the sequence :

$$f(x+\delta x, y+\delta y, t) = f(x,y,t) + a_t \delta x + b_t \delta y + n_1 \quad (5)$$

$$f(x+\delta x, y+\delta y, t+1) = f(x,y,t+1) + a_{t+1} \delta x + b_{t+1} \delta y + n_2 \quad (6)$$

where n_1 and n_2 are assumed to be independent zero-mean Gaussian noises with same variance σ^2.

The reliability of the constraint equation is tested by considering two competing hypotheses denoted H_0 and $\bar{H}_0$, where :

$$\begin{aligned} H_0 &: \{a_t = a_{t+1} \ and \ b_t = b_{t+1}\} \quad in \ W(s) \\ \bar{H}_0 &: \{a_t \neq a_{t+1} \ or \ b_t \neq b_{t+1}\} \quad in \ W(s) \end{aligned} \quad (7)$$

$W(s)$ designates a local window centered at point s, in which the parameters are estimated. The likelihood functions under each hypothesis are computed, assuming gaussian noises with same variances for n_1 and n_2 and the log-likelihood ratio is compared to a threshold. As in the case of the ME constraint, we define a binary validation factor for the motion constraint at site s :

$$\begin{aligned} \xi_g(s) &= 1 \quad if \ H_0 \ is \ selected \\ \xi_g(s) &= 0 \quad if \ H_0 \ is \ rejected \end{aligned} \quad (8)$$

This hypothesis test has some analogy with one of the likelihood tests described by Hsu *et al.* in [30] to detect temporal changes in image sequences (namely the one corresponding to the linear intensity model). Changes are recovered by detecting variations in the local parametrization of the brightness surface. These models can be either of constant, linear or quadratic type. Although the resulting techniques

are similar, the problem at hand in [30] yet remains different from ours.

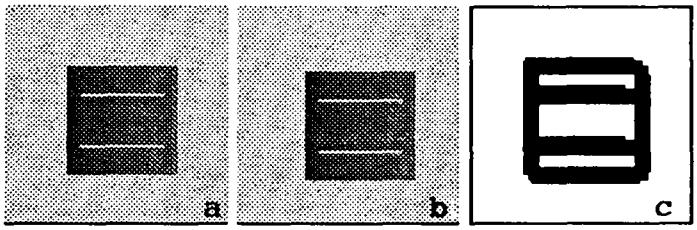


Figure 5 : binary confidence factor on the "square sequence".
a) Image t b) Image t+dt c) Confidence factor ξ_g.

Fig. 5 shows the sites corresponding to $\xi_g(s) = 0$ issued from test (7) on the *square sequence*. Let us notice that those regions are closely related to the occlusion areas (near the square boundaries) and to spatial discontinuities in the intensity pattern (along the two central lines). In these different areas the image flow constraint is indeed invalid : in the following, ξ_g will help us to disregard those wrong local constraints.

As intuitively expected comparing Fig. 5c and Fig. 3d, one can see that the two motion constraints are complementary in this example. Experiments with other sequences have given similar results : the gradient-based constraint is invalid in regions of "spatiotemporal discontinuity" (within occlusions, on sharp intensity edges and in highly textured regions) whereas the ME method reveals reliable at those locations (excepted in textured regions). This experimental statement justifies the use of those two particular motion constraints.

3 A global bayesian formulation for multimodal motion estimation

Global bayesian estimation defines a coherent mathematical framework to extract labels describing motion from image sequences.

The estimation process can be outlined as follows :

1. One or more specialized low-level modules extract from the image sequence features (gradients, moving edges...) that will be used as observations in the estimation process.

2. Observations are combined within local photometric and structural models with a priori generic knowledge on the expected result, in order to derive estimates of the unknown labels.

The spatial interactions between observation fields and motion labels are specified using Markov Random Field (MRF) models. In the MRF model designed here, local motion discontinuities are simultaneously estimated with the velocity field and multiple local constraints contribute to the estimation of those fields. Intensity edges are used as an additional evidence to support the estimation of motion boundaries (Fig. 6).

3.1 Observations and labels supporting motion information

In the estimation process, informations about motion are summarized in following labels :

- vectorial labels $\vec{\omega}_s$, $(\vec{\omega}_s \in \mathcal{R}^2)$ corresponding to the velocity field $\vec{\omega} = \{\vec{\omega}_s \ , \ s \in S\}$ where S denotes the set of pixel sites in the image plane. A local velocity vector is thus associated to every point s in the image plane.

- A set of discrete labels $\gamma = \{\gamma_d \ , \ d \in D\}$ describing local motion discontinuities. D denotes the set of edge sites located midway between the pixel sites. There are three possible states for motion discontinuities : $\gamma_d = 0, 1$ or -1.

  $\gamma_d = \pm 1$ (resp. $\gamma_d = 0$) means that a motion discontinuity (resp. no motion discontinuity) is present at location d. In case that $\gamma_d = \pm 1$, the sign of γ_d codes the relative position of the occluding surface with respect to the motion discontinuity, (see Fig. 7). $\gamma_d = +1$ (resp. $\gamma_d = -1$) means that for a vertical discontinuity the occluding region is on the right-hand side (resp. on

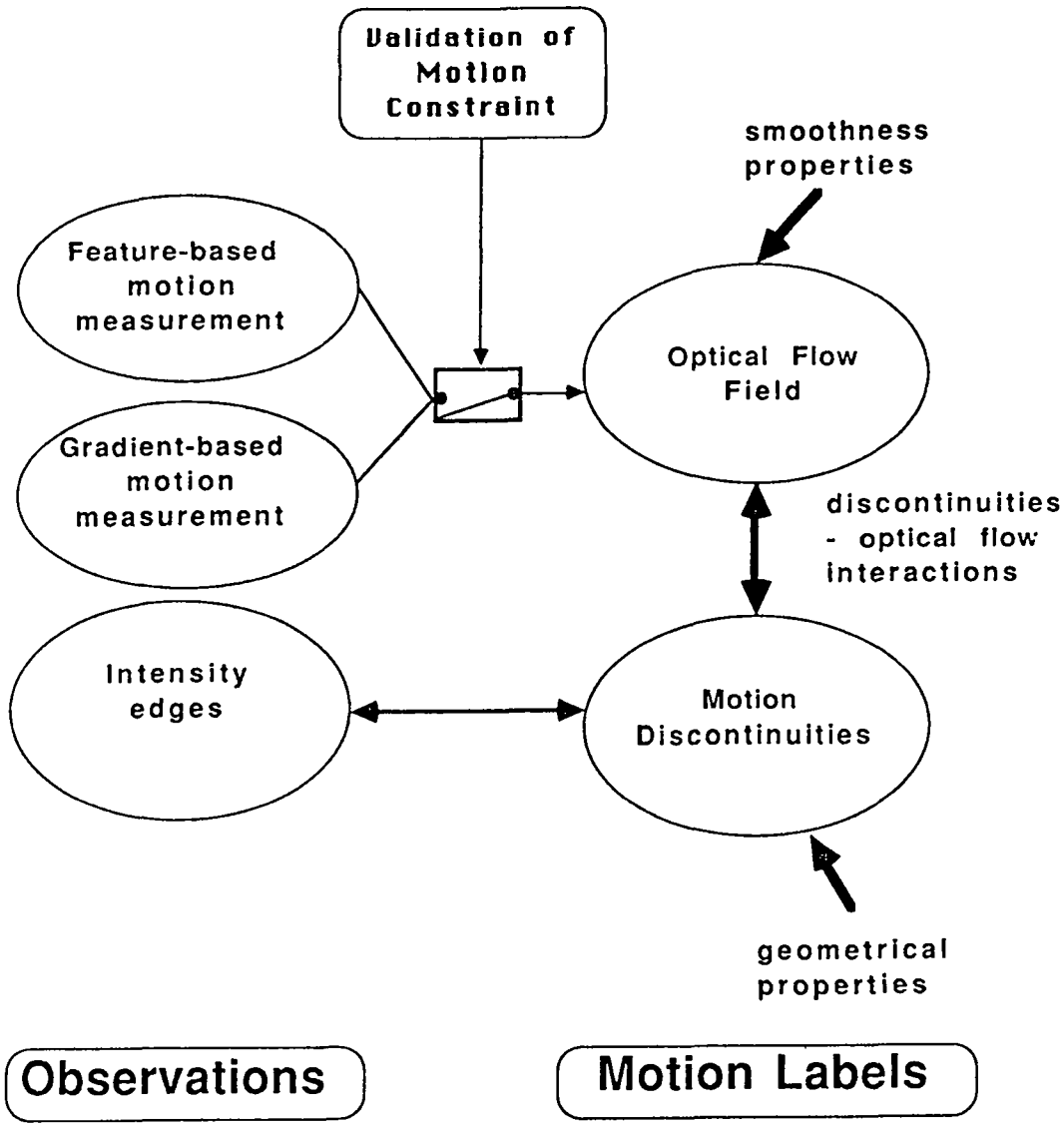


Figure 6 : The multimodal interaction model : interactions between motion labels and observations.
Thanks to the MRF model, the motion discontinuities are estimated jointly with the optical flow field. The motion measurement relies on two complementary motion constraints : gradient and feature-based. The reliability of local motion constraints is tested and they contribute to the global estimation only if they are valid. The estimation of motion discontinuities is supported by intensity edges. Geometrical constraints are also defined on the desired discontinuity configurations.

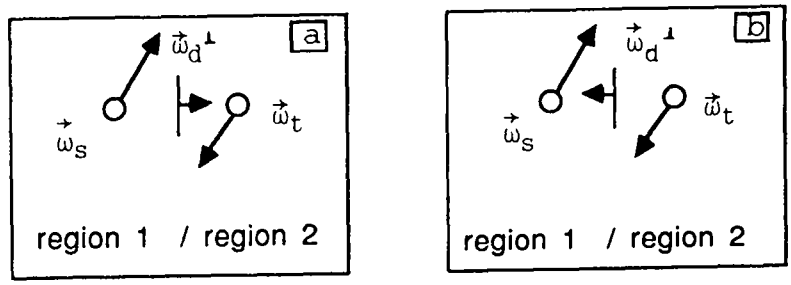


Figure 7 : The sign of label γ_d defines the region to which the occluding edge belongs (case of a vertical edge)
a) Case $\gamma_d = -1$: the occluding edge belongs to region 1 ($\vec{\omega}_d^{\perp}$ and $\vec{\omega}_s$ are consistent).
b) Case $\gamma_d = +1$: the occluding edge belongs to region 2 ($\vec{\omega}_d^{\perp}$ and $\vec{\omega}_t$ are consistent).

the left-hand side). A similar code is used for horizontal discontinuities. This three state edge description is used to propagate the moving edge constraint to the proper region (see Section 2.2), but can also be considered as a useful by-product of the estimation scheme.

Observations correspond to the output of four independent modules :

- a first module computing the spatial and temporal derivatives $\mathcal{D}f = \{\vec{\nabla} f(s), \frac{\partial f}{\partial t}(s),\ s \in S\}$ of the intensity function f at every vector site s.

- the moving edge estimation module described in Section 2.2, which yields displacement information about intensity edges located on grid D (only the displacements perpendicular to the edges are determined). The set of sites $d \in D$, for which a moving edge exists, will be denoted D_{me}. The corresponding local motion measurement set is denoted $\omega^{\perp} = \{\vec{\omega}_d^{\perp},\ d \in D_{me}\}$.

- The validation factors defined in Section 2.3, $\xi_g = \{\xi_g(s),\ s \in S\}$, $\xi_g(s) \in \{0,1\}$ and $\xi_{me} = \{\xi_{me}(d),\ d \in D\}$, $\xi_{me}(d) \in \{0,1\}$. Observations ξ_g state, at every pixel location s, whether the image flow constraint is reliable or not (the same holds for ξ_{me} at edge locations d for the moving edge constraint).

- A spatial intensity edge detector derived from Canny's criterion proposed by Deriche, [14] , provides a binary information about intensity discontinuities on sites $d \in D$. $\eta = \{\eta_d, d \in D\}$ designates the binary map output of the

intensity edge detector ($\eta_d = 1$ if an intensity edge is detected). Following [18] intensity edges will be used as partial evidence supporting the state of motion discontinuities γ_d at the same locations.

As explained in Section 2.2, the output of the intensity edge detector is also used in the moving edge estimator to reduce the global computation cost. The spatial locations of the moving edges $\omega^{\perp}$ thus coincide with the locations of the intensity edges used to support motion boundary detection.

3.2 Global bayesian decision and the MAP criterion

3.2.1 MAP criterion and the joint distribution

The Maximum A Posteriori (MAP) criterion has been widely used in the context of global bayesian decision, [7, 20, 35, 41] . To derive the unknown label fields $(\vec{\omega}, \gamma)$ from the observed fields $(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta)$, the following optimization problem has to be solved :

$$\max_{\{\vec{\omega},\gamma\}} \; p(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta \; , \; \vec{\omega}, \gamma) \tag{9}$$

where $p(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta \; , \; \vec{\omega}, \gamma)$ is the joint distribution of the observed and hidden variables.

Other decision criteria include the "Maximizer of the Posterior Marginal" (MPM), the "Tresholded Posterior Mean", [38] and the "Iterated Conditional Modes" (ICM), [4] . Computationally demanding stochastic relaxation algorithms are generally necessary to compute exact MAP solutions. Deterministic descent algorithms such as ICM or HCF (Highest Confidence First, [11]) can be use instead. They converge to a local maximum of the a posteriori distribution, but are far less cpu time consuming. ICM, [4] , is often used to compute suboptimal solutions when a good initial guess is available. In our case, we used the MAP criterion associated to such a deterministic optimization algorithm.

3.2.2 Modeling the joint likelihood using Markov Random Fields

The distribution of observations and motion labels are specified using a coupled Markov Random Field (MRF) model whose distribution is written in the following form ([20]) :

$$\begin{aligned} p(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta \; , \; \vec{\omega}, \gamma) &= \frac{1}{Z} \; exp \; - U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta \; , \; \vec{\omega}, \gamma) \\ \text{where :} \quad U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta \; , \; \vec{\omega}, \gamma) &= \sum_{c \in C} V_c(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta \; , \; \vec{\omega}, \gamma) \end{aligned} \tag{10}$$

is called the energy function of the MRF. The lowest energies correspond to the most likely configurations. Such a formulation is possible since we assume that the

interactions between the different variables remain local, with respect to a chosen neighborhood system ν (see [20] for a complete theory of MRF). C denotes the set of cliques associated to neigborhood system ν. Cliques c are subsets of sites which are mutual neighbors. The potential function V_c is locally defined on clique c and expresses the local interactions between the different variables of the clique. The form of the potential functions is of course problem dependent. The functions that we have defined for the motion estimation scheme integrate the different modeling aspects already discussed : regularization of the velocity field along with preservation of motion boundaries, multimodal cooperation between different measurement sources, disregard of invalid local motion constraints (in particular in the occlusion regions), processing of motion discontinuities... Within this framework, finding the Maximum A Posteriori estimate comes to minimizing the global energy function U.

The neighboring system ν is defined on sets S and D as explained in Fig. 8a. Interactions between observations, velocities, and motion discontinuities are supported by mixed cliques, whereas edge cliques support the geometric properties of motion discontinuities (Fig. 8b).

Interactions between variables are modeled through the following decomposition of the global energy function :

$$\begin{aligned} U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{me}, \eta\ ,\ \vec{\omega}, \gamma) \ &= \ U_1(\mathcal{D}f, \xi_g, \vec{\omega}) \ + \ U_2(\omega^{\perp}, \xi_{me}, \vec{\omega}, \gamma) \\ &+ \ U_3(\vec{\omega}, \gamma) \ + \ U_4(\eta, \gamma) \ + \ U_5(\gamma) \end{aligned}$$

Each term is itself decomposed into local potential functions defined on the different cliques (for the notations see Fig. 7 and Fig. 8) :

$$\begin{aligned} U_1(\mathcal{D}f, \xi_g, \vec{\omega}) \ &= \ \alpha_1 \sum_{s \in S} \xi_g(s)\ (\vec{\nabla} f(s).\vec{\omega}_s \ + \ f_t(s))^2 \\ U_2(\omega^{\perp}, \xi_{me}, \vec{\omega}, \gamma) \ &= \\ \alpha_2 \sum_{s \in S, d \in D_{me},\ (s,d) \in \{C_3\} \cup \{C_4\}} & \xi_{me}(d)\ (\vec{\omega}_s . \frac{\vec{\omega_d}^{\perp}}{\| \vec{\omega_d}^{\perp} \|} \ - \ \| \vec{\omega_d}^{\perp} \|)^2\ \frac{1}{2}(-|\gamma_d| - \gamma_d + 2) \\ + \ \alpha_2 \sum_{t \in S, d \in D_{me},\ (t,d) \in \{C_1\} \cup \{C_2\}} & \xi_{me}(d)\ (\vec{\omega}_t . \frac{\vec{\omega_d}^{\perp}}{\| \vec{\omega_d}^{\perp} \|} \ - \ \| \vec{\omega_d}^{\perp} \|)^2\ \frac{1}{2}(-|\gamma_d| + \gamma_d + 2) \\ U_3(\vec{\omega}, \gamma) \ &= \ \sum_{(s,t,d) \in \{C_5\} \cup \{C_6\}} \Phi_{\alpha_3}(\| \vec{\omega}_s \ - \ \vec{\omega}_t \|)\ (1 - |\gamma_d|) \\ U_4(\eta, \gamma) \ &= \ \alpha_4 \sum_{d \in D} (1 - \eta_d)|\gamma_d| \\ U_5(\gamma) \ &= \ \sum_{c \in \{C_7\} \cup \{C_8\}} V_c(\gamma) \end{aligned} \tag{11}$$

where Φ_{α_3} is given by :

$$\Phi_{\alpha_3}(\| \vec{\omega}_s - \vec{\omega}_t \|) = sign(\| \vec{\omega}_s - \vec{\omega}_t \| - \alpha_3) \frac{(\| \vec{\omega}_s - \vec{\omega}_t \| - \alpha_3)^2}{{\alpha_3}^2}$$

the α_i, $i = 1,..4$, are the model parameters
and the C_i, $i = 1,..8$ denote the different clique types depicted in Fig. 8b.

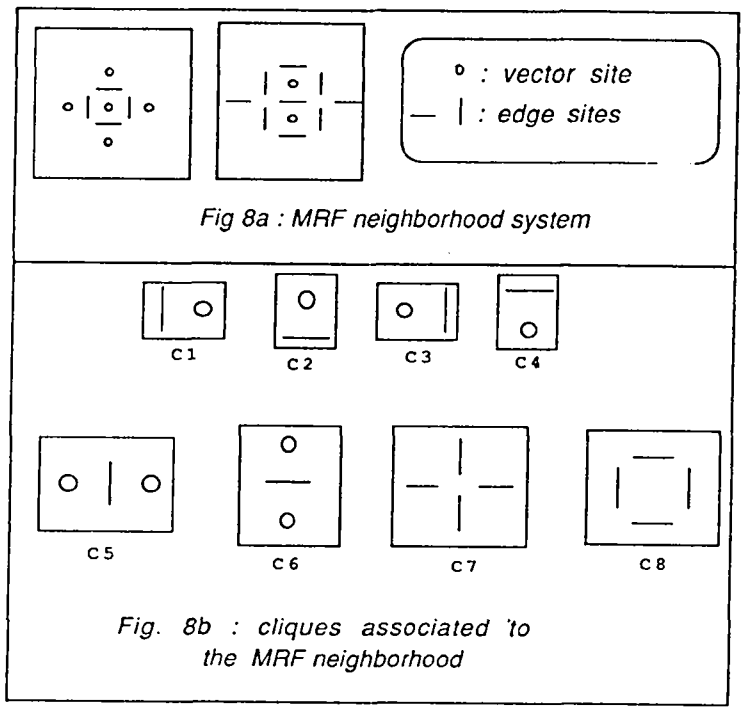


Figure 8 : MRF neighborhood system including vector and edge sites.

In this decomposition each term expresses a different interaction model, which brings its partial contribution to the global estimation process. The different terms can be interpreted as follows.

Energies U_1 and U_2 (motion measurement constraints)

Energies U_1 and U_2 are related to the motion constraint equations (2) and (3) upon which the motion estimation is based. Those energy functions take into account the confidence factors discussed in Section 2.3. For sites s belonging to smooth spatiotemporal regions ($\xi_g(s) = 1$), the image flow constraint equation is applied (energy U_1), whereas in sites where $\xi_{me}(s) = 1$, (i.e. presence of a moving edge), the moving edge constraint is considered (energy U_2).

If the moving edge at site d corresponds to a motion discontinuity $|\gamma_d| = 1$, the moving edge constraint is only propagated to the proper region depending on the

sign of γ_d. For instance for a vertical motion boundary (see Fig. 7), when $\gamma_d = 1$, $\frac{1}{2}(-|\gamma_d| - \gamma_d + 2) = 0$ and $\frac{1}{2}(-|\gamma_d| + \gamma_d + 2) = 1$; hence, the moving edge constraint is only propagated to the vector $\vec{\omega}_t$ belonging to the right-hand side region (Fig. 7). When $\gamma_d = 0$, i.e. when there is no motion discontinuity at site d, the moving edge constraint is propagated to the regions on both sides.

Conversely when the velocities are given, two configurations may be encountered :
- the measured orthogonal component $\vec{\omega_d}^{\perp}$ is exactly consistent with the neighboring velocities $\vec{\omega_s}$ and $\vec{\omega_t}$, that is $\vec{\omega_t}.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \vec{\omega_d}^{\perp} = \vec{\omega_s}.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \vec{\omega_d}^{\perp} = 0$. In this case, energy U_2 has no effect on the choice of γ_d.
- a discrepancy exists between $\vec{\omega_d}^{\perp}$ and its neighbor vectors. In this case, energy U_2 favours either value $\gamma_d = -1$ or $\gamma_d = +1$ according to the lowest local energy. For a vertical discontinuity for instance, if the perpendicular component $\vec{\omega_d}^{\perp}$ is consistent with vector $\vec{\omega}_s$ (Fig. 7), the value -1 is assigned to γ_d. This means that the occluding surface is on the left-hand side of the boundary in this case. The occluding region is the one containing the velocity vector ($\vec{\omega}_s$ or $\vec{\omega}_t$) the most consistent with the observed component $\vec{\omega_d}^{\perp}$.

Let us point out that energy function U_2 does not decide between states $\gamma_d = 0$ and $|\gamma_d| = 1$, i.e. whether a motion boundary has to be introduced or not. Energy U_2 only makes a selection among states $\gamma_d = 1$ and $\gamma_d = -1$, that is, assigns an existing motion boundary to the region which it belongs to. The decision for the placement of a motion boudary at a given location is inferred from energy term U_3, based on the variations in the estimated velocity field.

Energy U_3 (velocity field smoothing while preserving discontinuities)

Interactions between velocities and motion discontinuities are supported by mixed cliques (s, t, d). The chosen potential function smooths out the motion field using terms of the form $\| \vec{\omega}_s - \vec{\omega}_t \|$, corresponding to first order derivatives (second order terms were also used, but appeared more sensitive to noise). Velocity smoothing is inhibited when a motion boundary is present ($|\gamma_d| = 1$). Conversely when the vector field shows important variations in edge vicinity ($\| \vec{\omega}_s - \vec{\omega}_t \| > \alpha_3$), the placement of a motion discontinuity at that site is favored ($|\gamma_d| = 1$).

Energy U_4 (interactions between motion discontinuities and intensity edges)

Intensity edges are used as partial evidence for the determination of motion boundaries. In real world scenes indeed, a 3D configuration resulting in a motion discontinuity generally also contributes to an intensity edge. Hence, following [18] , we assume that motion discontinuities have low probability to appear when there is no intensity edge at the same location. This is specified using energy function U_4 with a large positive value for parameter α_4. U_4 prevents configurations in which

$|\gamma_d| = 1$ and $\eta_d = 0$ from appearing.

Energy U_5 (edge geometry)

The edge cliques considered in U_5 help to discourage undesirable geometric configurations (edge ending, isolated or double edges [20]). Two main methods for specifying the geometrical properties of edges in MRF have been proposed, [19, 20] . The first approach, [20] , consists in assigning different weights to the different possible local edge configurations defined on the geometrical cliques. A large weighting on a configuration tends to discourage this configuration, [20] . The main drawback of this method is to introduce an important number of additional parameters in the model, corresponding to the different weights. The tuning or learning of those parameters is generally not an easy task. As we are concerned with a three state discontinuity description and with the chosen neighborhoods, the number of local configurations is as large as 162. Therefore we have resorted to an alternate approach recently described in [19] . This other solution consists in introducing *forbidden* edge configurations. Each forbidden local configuration (edge endings, isolated or double edges, impossible configurations of occluding/occluded regions), weighs an elementary weight of 1 in energy U_5. The following constrained optimization problem is then solved :

$$\min_{\{\vec{\omega},\gamma\}} (U_1 + U_2 + U_3 + U_4) + \alpha_5 U_5 \quad \text{with} \quad \alpha_5 = +\infty \tag{12}$$

A constrained optimization may be obtained in practice by letting $\alpha_5 \nearrow +\infty$ (see [19] for convergence theorems).

3.3 Energy minimization using a deterministic relaxation

Finding the MAP estimate of the fields $\vec{\omega}$ and γ is equivalent to minimizing the global energy function $U(\mathcal{D}f, \omega^{\perp}, \xi_g, \xi_{mc}, \eta\ ,\ \vec{\omega}, \gamma)$. This global energy function depends both upon continuous and discrete valued variables ($\vec{\omega}$ and γ). To reach configurations close to the global minimum of an energy, stochastic optimization methods for continuous and discrete variables have been studied [21, 39] . As recently pointed out by Woods *et al.*, [33] the convergence of simulated annealing has been proven only in the case of compact configuration spaces. In our case the magnitudes of the optical flow vectors are not bounded *a priori* (even though reasonable constraints could be imposed on the largest possible magnitude). Moreover, the stochastic optimization algorithms are very time consuming, especially for continuous variables. Most of the recent papers resort to deterministic schemes which are more appealing, as far as computation time is concerned. Deterministic relaxation converges to a local minimum of the energy function, but this loss of optimality may be compensated by an appropriate initial guess. Besides, in many cases the

suboptimal solution can be considered as a "sensible" solution. In our experiments, we use a modified version of Iterated Conditional Modes [4] , a deterministic alternative to simulated annealing. In this relaxation scheme the final result depends on initialization and site visiting order. Satisfactory results are obtained by initializing vectors with $\vec{0}$ and motion boundaries with the state of intensity edges η_d. This proves that for the optimization problem at hand, the initial guess for the velocity field is not so critical.

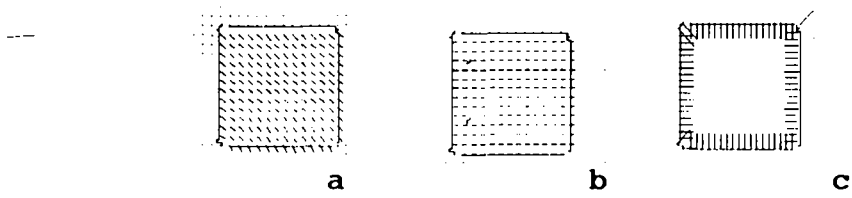


Figure 9 : velocity fields and motion discontinuities obtained on the square sequence.

a) Optical flow and motion discontinuities obtained by the multimodal estimation scheme.

b) Fields obtained without the moving edge constraint ($\alpha_2 = 0$).

c) Differenciation of occluding surface and occluded surface. Each symbolic vector points to the interior of the region to which the discontinuity belongs.

The result of the multimodal method on the square sequence (Fig. 1a) is given in Fig. 9. (The confidence factors associated to the different motion constraints are shown in Fig. 3d and 5c). The designed interaction model captures the real motion boundaries and excellent accuracy is reached, especially near discontinuities as it can be seen. In this example the estimation process propagates the moving edge constraint from the two central intensity lines and the square boundaries to the rest of the field. The sign of γ_d expresses, for every motion discontinuity, the relative position of occluding and occluded surfaces (Fig. 9c). Each symbolic dash in Fig. 9c points to the interior of the region to which the motion discontinuity belongs. For comparison Fig. 9b presents a motion field obtained with a modified energy function which does not include the moving edge constraint. One can verify that only the horizontal component of the velocity field could be recovered in this case, since the gray value function in the square remains constant along the vertical direction. On the other hand, the cooperation with the moving edge constraint allows to recover the vertical component (see Fig. 9a).

3.4 Description of the implemented algorithm

As explained above, a deterministic optimization algorithm has been used. The algorithm is a modified version of ICM (for "Iterated Conditional Modes"), [4] . It has been adapted to handle continuous-valued variables and to allow an optimization under geometrical constraints. In such a scheme, the global energy function (Equ. (11)) is minimized by sequentially updating the different sites of the velocity and discontinuity fields. The site visiting order is raster scan, with reverse order after every full sweep of the image. At a given location, the label value assigned to a site is the one maximizing the decrease of the global energy function. Thanks to the decomposition of the MRF into local interaction terms, it turns out that the relaxation scheme only involves the visited site and its neighbors, [20] . Vector sites and discontinuity sites are visited in turn. The expression of the local energy functions that are minimized at every site s and d are derived from equations 11. However, to make the computation easier we consider some simplifications in those local energy functions :

- Local energy for velocity vector $\vec{\omega}_s$:

$$\begin{aligned} E(\vec{\omega}_s) &= \alpha_1\, \xi_g(s)\, (\vec{\nabla} f(s).\vec{\omega}_s + f_t(s))^2 \\ &+ \alpha_2 \sum_{d|\ (s,d)\in\{C_3\}\cup\{C_4\}} \xi_{me}(d)\, (\vec{\omega}_s.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \|\vec{\omega_d}^{\perp}\|)^2\, \frac{1}{2}(-|\gamma_d| - \gamma_d + 2) \\ &+ \alpha_2 \sum_{d|\ (s,d)\in\{C_1\}\cup\{C_2\}} \xi_{me}(d)\, (\vec{\omega}_s.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \|\vec{\omega_d}^{\perp}\|)^2\, \frac{1}{2}(-|\gamma_d| + \gamma_d + 2) \\ &+ \alpha_3{}' \sum_{(t,d)|\ (s,t,d)\in\{C_5\}\cup\{C_6\}} \|\vec{\omega}_s - \vec{\omega}_t\|^2\, (1 - |\gamma_d|) \end{aligned} \tag{13}$$

- Local energy for discontinuity γ_d :

$$\begin{aligned} E(\gamma_d) &= \alpha_2\, \xi_{me}(d)\, (\vec{\omega}_s.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \|\vec{\omega_d}^{\perp}\|)^2\, \frac{1}{2}(-|\gamma_d| - \gamma_d + 2) \\ &+ \alpha_2\, \xi_{me}(d)\, (\vec{\omega}_t.\frac{\vec{\omega_d}^{\perp}}{\|\vec{\omega_d}^{\perp}\|} - \|\vec{\omega_d}^{\perp}\|)^2\, \frac{1}{2}(-|\gamma_d| + \gamma_d + 2) \\ &+ \Phi_{\alpha_3{}''}(\|\vec{\omega}_s - \vec{\omega}_t\|)\, (1 - |\gamma_d|) \\ &+ \alpha_4\, (1 - \eta_d)|\gamma_d| \\ &+ \alpha_5 \sum_{\{C_7\},\{C_8\}\ |\ d\ \in\ \{C_7\},\{C_8\}} V_c(\gamma) \end{aligned} \tag{14}$$

and $\Phi_{\alpha_3{}''}$ is given by :

$$\Phi_{\alpha_3{}''}(\|\vec{\omega}_s - \vec{\omega}_t\|) = sign(\|\vec{\omega}_s - \vec{\omega}_t\| - \alpha_3{}'')\, \frac{(\|\vec{\omega}_s - \vec{\omega}_t\| - \alpha_3{}'')^2}{\alpha_3{}''^2}$$

The terms of the local energies are derived from the global one (equation 11), apart from one exception concerning the smoothing term $\Phi_{\alpha_3}(\| \vec{\omega_s} - \vec{\omega_t} \|)$ which has been replaced by the simplified quadratic form $\alpha_3' \| \vec{\omega_s} - \vec{\omega_t} \|^2$ in the local energy function $E(\vec{\omega_s})$. Using such a simplified form makes the computation of the local minimum of $E(\vec{\omega_s})$ easier since it becomes quadratic with respect to $\vec{\omega_s}$. No visible degradation on the final velocity fields was noticed when this simplification was used. The results presented in Section 4 and 5 have been computed with these energy functions (equations 13 and 14).

Following [19] , the optimization is constrained by letting parameter α_5 (Equ. 12) carefully grow to $+\infty$. The rate of increase of α_5 is chosen not to get stuck too quickly into a local minimum.

If n designates the *iteration* number (an *iteration* is here a full sweep on the image), following schedule has been adopted, [19] :

$$\alpha_5 = \alpha \, log \, n \quad with \quad \alpha = \frac{4}{{\alpha_3''}^2} \tag{15}$$

The other model parameters are either set to a fixed value, whatever the sequence at hand, or computed "on line" : $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_4 = \frac{4}{{\alpha_3''}^2}$. Indeed, the only parameters that need to be tuned are α_3' and α_3'' which weight the interactions between velocities and motion discontinuities. α_3' and α_3'' have been chosen by trial and error for the different sequences and may change from one sequence to another. Clearly an efficient data driven parameter identification method would be of great interest here.

As the relaxation is deterministic, the final solution *a priori* depends on initialization and site visiting order. As we have seen, the initial guess for the velocity field is not critical (and the velocity vectors are initialized to $\vec{0}$). Conversely, the motion boundaries reveal more sensitive to initial conditions and visiting order. In particular, if the discontinuity sites are updated from the beginning of the optimization, it is observed that the process gets stuck after a few iterations into an undesired local minimum. This has led us to conceive a two-step optimization algorithm that can be described as follows :

- *initialization* : $\vec{\omega_s} = \vec{0}$, $s \in S$ and $\gamma_d = \eta_d$, $d \in D$
  i.e. the velocities are initialized by $\vec{0}$ and the motion boundaries with the state of the intensity edges.

- *first step* (n_1 iterations) : the motion boundaries are frozen to the initial state (i.e. the state of intensity edges). The relaxation is performed until convergence is reached on the velocity field. As a matter of fact, the relaxation is stopped as soon as the maximum relative variation on the norm of the

velocity vectors (between two successive iterations) goes below some specified threshold t ($t = 10^{-2}$ was taken in all experiments).

- *second step* (n_2 iterations) : using the fields estimated in step 1 as an initialization, the full relaxation (both on motion boundaries and velocity vectors) is pursued until convergence to the final estimates (the convergence criterion is the same as in step 1). The velocity and discontinuity sites are visited in turn.

In this strategy the determination of motion boundaries is undertaken only once the velocities are stabilized within the regions delineated by the intensity edges. Other strategies were tried (introducing initial velocity fields obtained by a local optimization technique assuming constant velocity in a small window, for instance), but the results were less satisfactory.

In our experiments, between 200 and 400 complete scans on the image are usually necessary to lead to configurations close to convergence. This number of iterations seems unusually large for a deterministic scheme. This is certainly due to the fact that the velocity vectors are continuous-valued : in discrete problems, convergence is generally faster (within 10 iterations for binary-valued fields, for example). The number of iterations can yet be reduced efficiently by using some extensions described in Section 5.

4 Experiments

Experiments have been carried out on several synthetic and real world sequences. We have focused here on five sequences corresponding to different classes of images and movements : indoor scenes and outdoor scenes, situations comprising static camera and moving objects and situations involving both camera and object motions. These sequences have been chosen to illustrate different contributions of the multimodal estimation process : preservation of motion boundaries, multimodal cooperation between different measurement sources, processing of occlusions...

4.1 Experiments on synthetic sequences

The contributions of the different modeling features are best highlighted on a synthetic sequence, in which motion can be controlled. The *moving square sequence* example (Fig. 9) has already examplified that, even for the very ambiguous intensity pattern considered in that case, the multimodal motion estimation scheme reconstructs an optical flow field close to the true motion both on a qualitative and quantitative point of view.

We present here a second synthetic sequence called *moving disks*, including two disks undergoing different motions and occluding each other (Fig. 10). The fore-

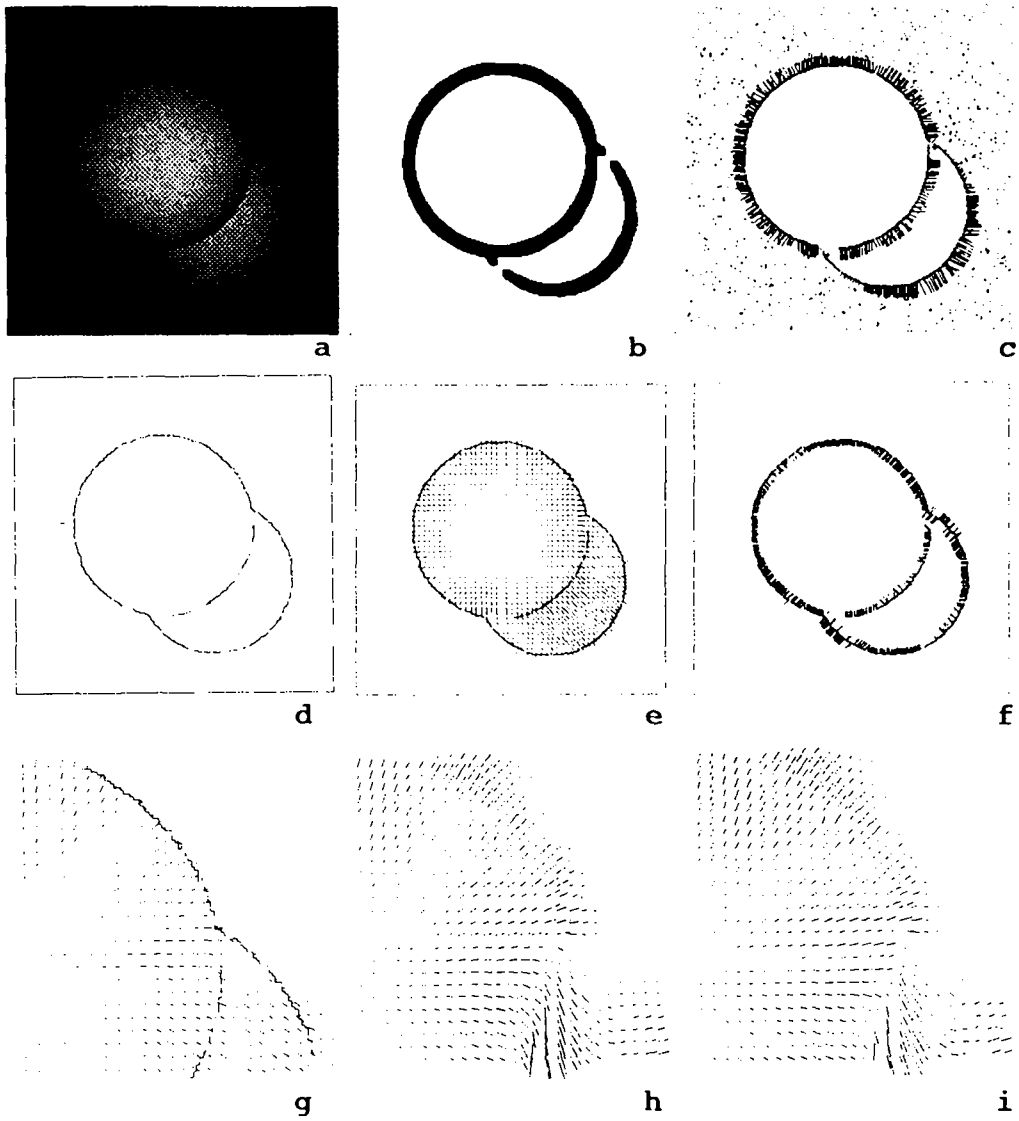


Figure 10 : motion estimation on the moving disks *sequence.*

a) First frame from the original sequence (256x256). The foreground disk undergoes a dilatation, the background disk a translation. White noise has been added to the background. b) Confidence factor for the gradient-based equation. The black areas corresponding to $\xi_g(s) = 0$ show the different occlusions in the scene. c) Intensity edges detected on the first frame by Deriche's edge detector, [14] . The perpendicular velocity displacements computed by the moving edge estimator, have been superimposed on the intensity edges. d) Motion boundaries estimated by the multimodal scheme (corresponding to $|\gamma_d(s)| = 1$). e) Resulting optical flow field (183 iterations, $\alpha_3' = 200$, $\alpha_3'' = 0.1$) f) In this figure each dash, computed from the sign of γ_d, points to the inner part of the occluding region. The disks are occluding the background. The occluded regions correspond to the background and the overlapping area between the two moving disks. g) Upper right part of Fig. 10e) (100x100). h) Result of Horn and Schunck's method. This field can be compared with the result of the multimodal scheme (Fig. 10g). i) Result obtained without taking the occlusions into account but handling the motion boundaries. This shows that a major part of the estimation error comes from the occlusion areas.

ground disk moves parallel to the axis of view, hence a dilatation is observed. This disk partially occludes another disk undergoing a translation of $3\sqrt{2}$ pixels towards the lower right corner. White noise has been added to the background.

The validation factors concerning the gradient-based motion constraint are derived from hypotheses test (7). Sites with validation factors equal to 0 are shown in Fig. 10b. One can notice that they are closely related to the different occlusion areas in the scene : part of the background covered by the two moving objects and partial overlapping of the two disks.

Fig. 10c depicts both the intensity edges detected on the original grey-level image and the perpendicular velocity components obtained from the ME estimator (see Section 2.2). The intensity edges are somewhat noisy, due to white noise added to the background. The velocity components in Fig. 10c measure, as expected, the displacement of the moving disks perpendicularly to the intensity edges.

The fields computed after the two-step relaxation process are drawn in Fig. 10d and 10e. 183 iterations were necessary here to reach convergence (model parameters : $\alpha_3' = 200$, $\alpha_3'' = 0.1$). Fig 10d presents the motion discontinuities corresponding to $|\gamma_d| = 1$. The proposed Markov interaction model filters the noisy detections on the background and only captures the true motion boundaries. Moreover, the sign of the motion discontinuity labels γ_d allows to differenciate the occluding regions from the occluded one. In Fig. 10f each dash points to the inner part of the occluding region. The performance of the method in the overlapping area between the two moving objects should be noticed : the dashes point to the true occluding disk. The result is the right one, except for two small parts of the boundary of the second disk which in fact slide parallel to themselves. There the local information remains ambiguous and it is not possible to differenciate the occluding region from the occluded one. Fig. 10e contains the estimated motion field : the accuracy is very satisfactory, when compared with the theoretical values, especially near the motion boundaries.

These results have been compared to the early gradient-based method proposed by Horn and Schunck, [29] , which takes into account neither motion discontinuities, nor occlusion areas. Parts of the fields obtained using the different methods appear in Fig 10g, 10h and 10i. Fig. 10g corresponds to the upper right part of the field estimated by our multimodal method (Fig. 10e). Fig. 10h presents the results of Horn and Schunck's method. As expected, the resulting motion field is blurred across the motion discontinuities. *However* introducing motion discontinuities without processing occlusions is not sufficient, as shown in Fig.10i. The field in Fig. 10i is computed while handling local motion boundaries, but without considering the invalid sites of Fig.10b. Let us recall that these sites correspond to the occlusion areas. As a result, the final field is very corrupted in this region. By comparing Fig. 10h and 10i one can conclude here that the major error source comes from the occlusion area rather than from an oversmoothing of the velocity field. This demonstrates that a specific processing of regions containing invalid observations is

a real contribution of the multimodal scheme.

4.2 Experiments on real world sequences

Several real-world sequences have been processed, which can be related to different contexts : traffic scene, TV sequences, etc...

- The first sequence called *parking lot* involves two cars manoeuvring in a park (Fig. 11). The moving vehicles correspond to the two cars in the foreground. This sequence can be considered as difficult because, unlike the previous examples, the camera also moves. The camera pans the scene, creating approximately a translation component in the image plane. This sequence contains textured areas, especially on the background trees.

- The second sequence *houses* has been acquired by panning an urban scene. Prominent grey level features appear on the houses along with quite uniform regions (Fig. 12).

- The last sequence called *interview* consists of a TV sequence : the woman on the right moves up and the camera follows her motion (Fig. 13).

The results presented here are computed from two successive frames out of the original sequences.

The result for the *parking lot* sequence is reported in Fig. 11c and 11d. The parameter values are : $n1 + n2 = 400$ *iterations*, $\alpha_3' = 1000.$, $\alpha_3'' = 0.02$. This sequence gives an example of the contribution of motion discontinuity processing to the estimation scheme. The estimated motion field fits with accuracy visual motion on the background trees induced by the camera panning, as well as the motion corresponding to the two manoeuvring cars (Fig. 11d). The optical flow on the road surface in the foreground is zero since the surface is almost completly uniform.

The local motion discontinuities reveal the silhouette of the two moving cars (Fig. 11c). Fig. 11b shows the intensity edges detected in the original grey-level image. Intensity edges are used as a support for the estimation of motion boundaries. The Markov interaction model filters the edges detected in the textured background and almost only captures the real motion discontinuities among the original intensity edges. This is an important feature of the algorithm, which allows to smooth the velocity field on regions of significant size, while preserving motion discontinuities. The motion boundaries can be considered as a by-product of the method, and provide useful cues for further interpretation of the dynamic content of the scene, for instance in applications such as object tracking or obstacle avoidance.

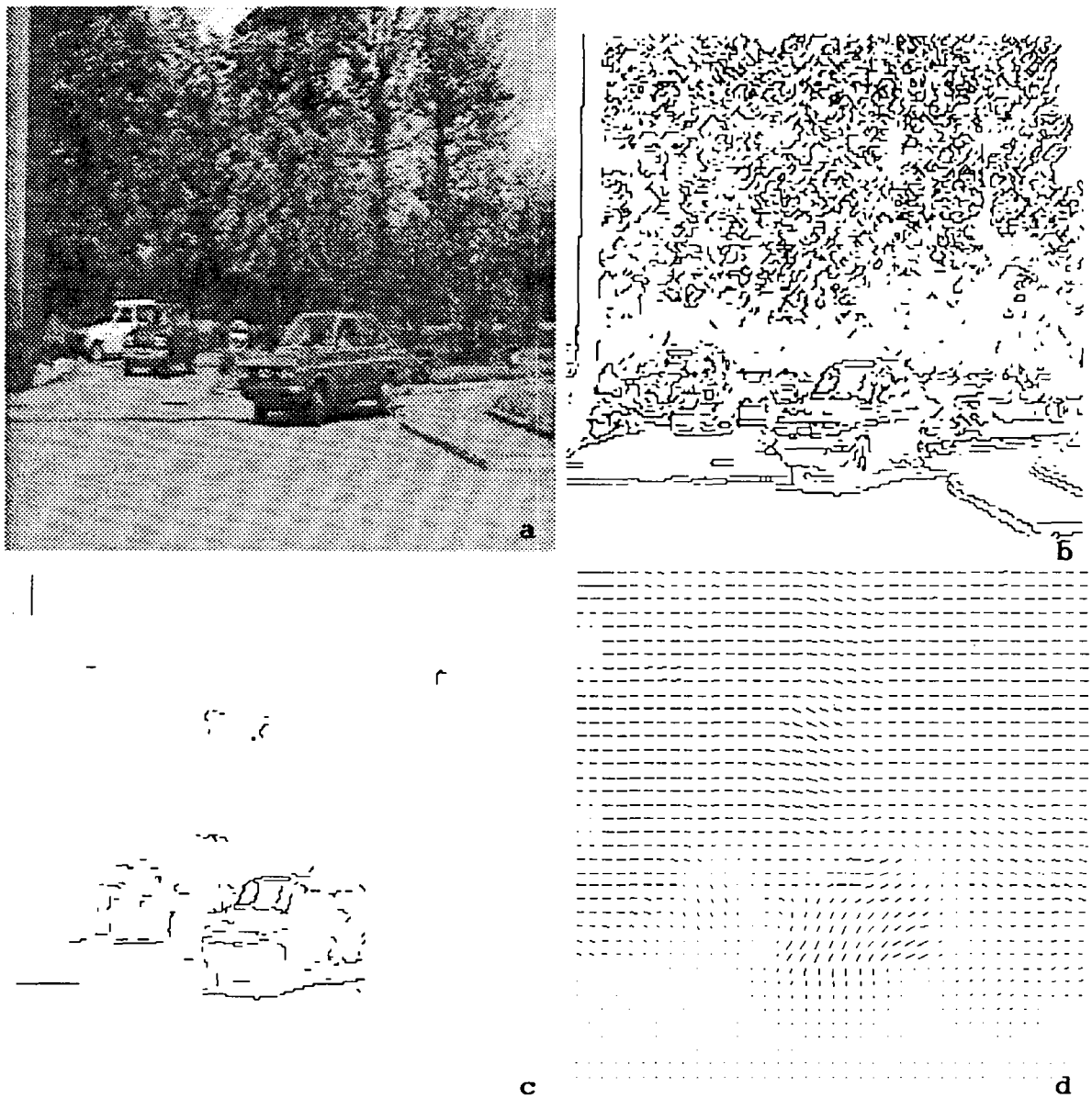


Figure 11 : Motion estimation and segmentation : "parking lot sequence" (by courtesy of Thomson-LER).

a) First frame of the sequence (224x224)

b) Intensity edges extracted from Fig. 11a

c) Estimated motion discontinuities (400 iterations, $\alpha_3' = 1000.$, $\alpha_3'' = 0.02$)

d) Associated velocity field (horizontally and vertically subsampled by 6) .

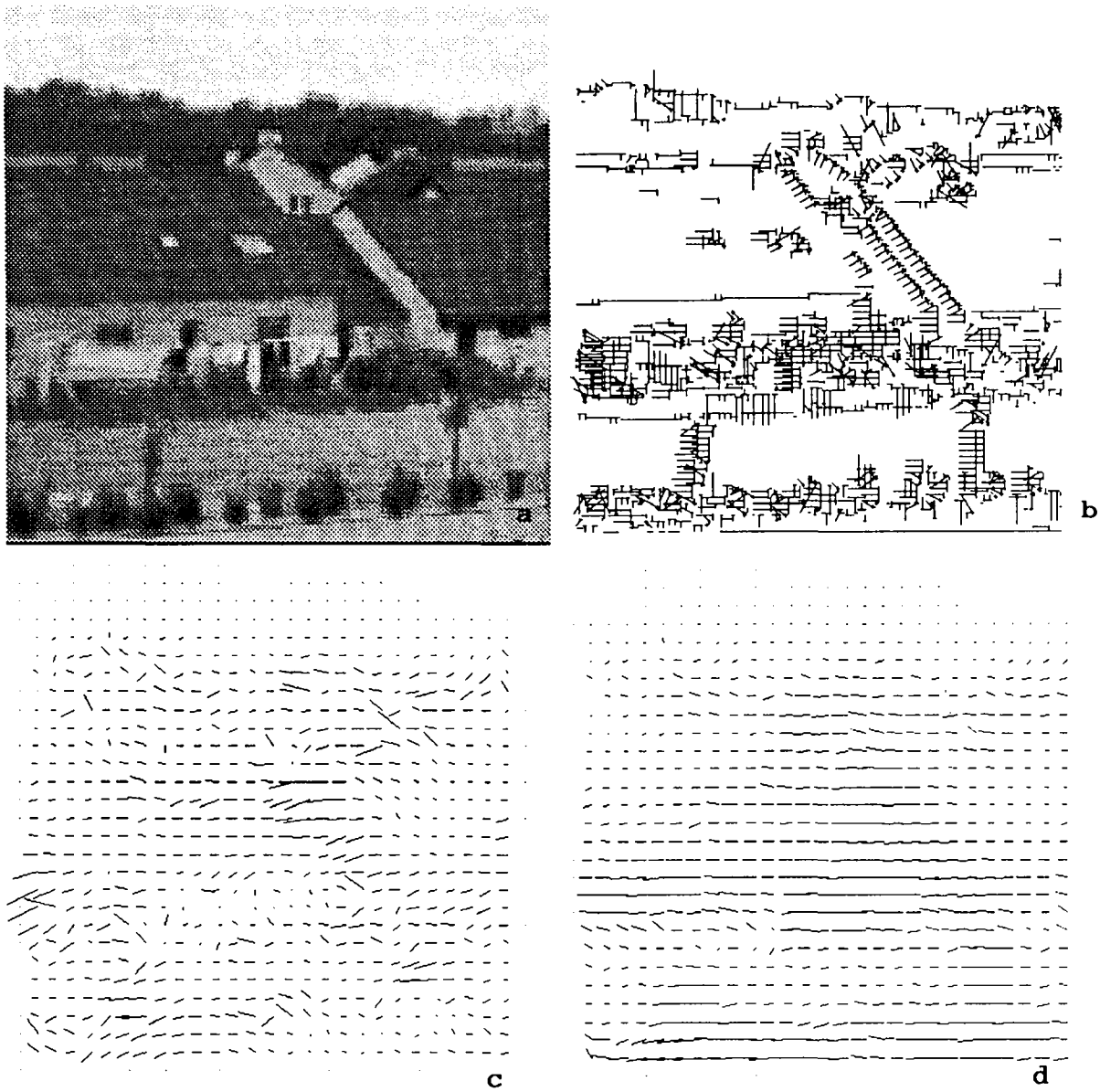


Figure 12 : Multimodal motion estimation : "urban scene" (by courtesy of Thomson-LER).

a) First frame of the sequence (170x170)

b) Perpendicular velocity components estimated on the intensity edges

c) Gradient-based only optical flow estimation (400 iterations, $\alpha_3' = 10.$, $\alpha_3'' = 1.$).

d) Multimodal optical flow estimation (horizontally and vertically subsampled by 6)(same parameter values as in Fig. 12 c).

The sequence *houses* (Fig. 12) was chosen to examplify the contribution of the moving edge constraint, when there are many regions of uniform intensity in the image. This is the case in the almost uniform areas of the house roofs and walls in Fig. 12a. Fig. 12b shows moving edges detected by the algorithm described in [6] . Fig. 12c and 12d present the resulting velocity field computed using two different methods. The first one only makes use of the gradient-based constraint, the second one includes the moving edge constraint. The improvement due to the multimodal cooperation scheme is quite visible (Fig. 12d). The visual motion corresponding to a translation in the image plane is well estimated in Fig. 12d, whereas it remains noisy in Fig. 12c. The image flow equation here does not bring sufficient information : the use of an additional constraint largely improves the result.

As far as the *interview* sequence is concerned, motion boundaries are closely related to the woman's movement (compare Fig. 13b with Fig. 13c). The estimated motion field accurately reproduces the visual motion in the background due to the camera panning. The complex motion of the woman moving up (Fig. 13d) is recovered with a good accuracy, especially near motion boundaries. It can be compared to the field computed from the Horn and Schunck method, Fig. 13f. The oversmoothing is very perceptible in this last case.

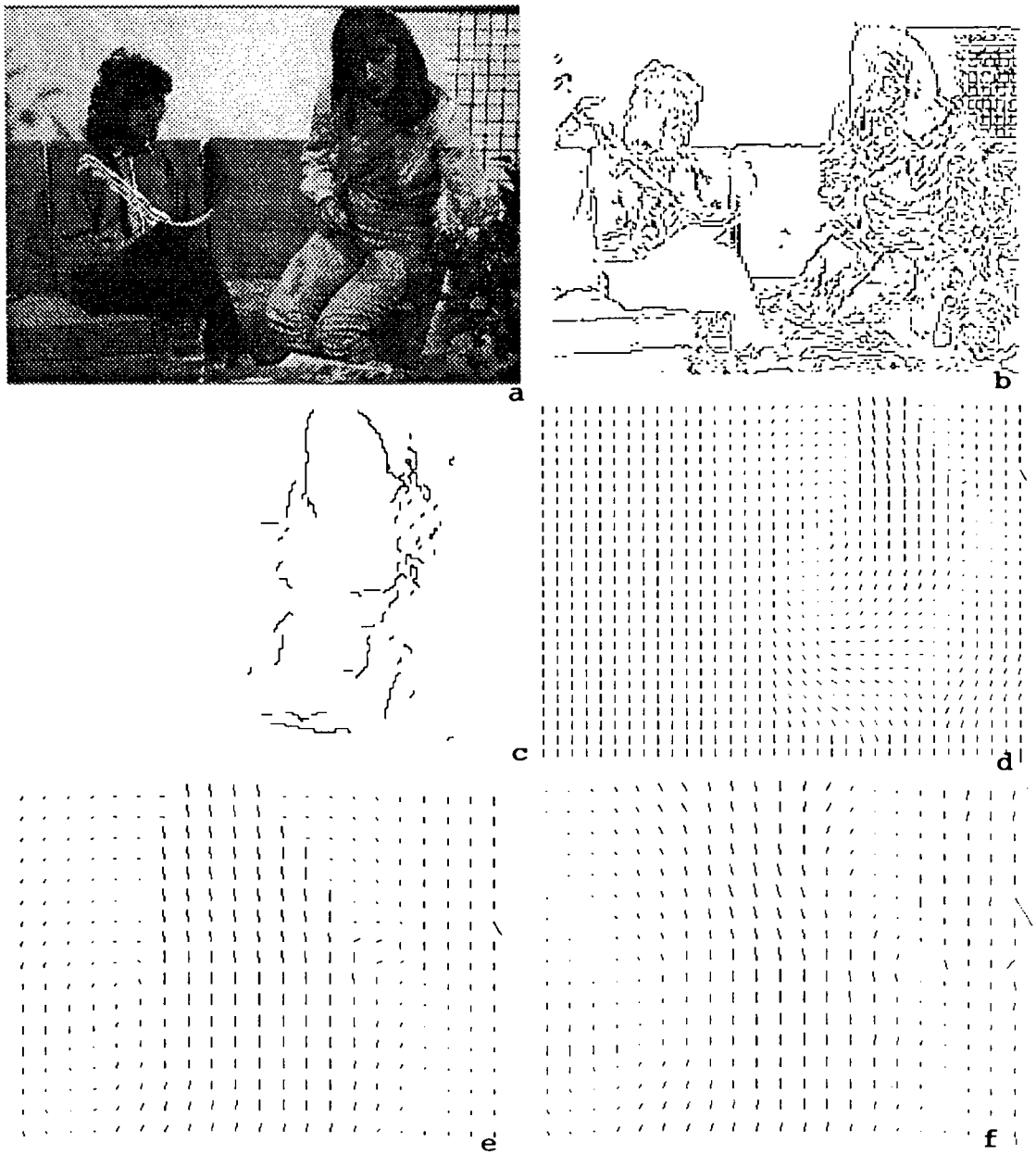


Figure 13 : Multimodal motion estimation : "Interview" sequence (by courtesy of BBC-UK).

a) First frame of the original sequence (134x168) : the woman at the right moves up and the camera follows her motion.

b) Intensity edges extracted from Fig. 13a

c) Motion boundaries estimated by the multimodal estimation scheme (400 iterations, $\alpha_3' = 1000.$, $\alpha_3'' = 0.15$*).*

d) Associated optical flow estimation (horizontally and vertically subsampled by 5).

e) Detail of the optical flow field of Fig. 13d showing the woman's head.

f) Result of Horn and Schunck method on the same detail as in Fig. 13e. The smoothing of the optical flow field across the discontinuities is visible.

5 Extensions towards multiresolution and multiframe motion analysis

As explained in Section 2.1, large displacements are generally not reachable using gradient-based motion estimation methods. This also concerns our multimodal scheme, which makes use of the gradient-based constraint. As soon as the displacements become large, the confidence factor associated to the gradient-based measurements decreases very quickly. As a consequence, a large part of the gradient-based measurements do not take part to the estimation of the final velocity field and in many cases there is not enough information available to get reasonable results. A classical solution to this problem is to use a multiscale image analysis. The large displacement vectors are determined at lower resolutions, where the interactions between spatial and temporal derivatives are maintained.

A second possible extension of the proposed scheme is related to the processing of multiple frames. Intrinsically, the method only considers two successive frames out of the sequence. There are no connections between the estimates derived at different times. However, the velocities usually vary smoothly along a sequence; hence the estimated velocities at time t can be used as an initialization for the estimation at time $t + dt$.

Those two extensions are of course classical ones : multigrid algorithms have been studied by Terzopoulos, [50] and Enkelmann, [16] for the estimation of optical flow fields. The multigrid methods have been adapted to relaxation algorithms for MRF models by Barnard, [3] in stereo matching and Konrad *et al.*, [35] for motion estimation. In this paper we resort to such an approach, close to the one reported in [16] or [35] . We describe it briefly in this section.

5.1 Multigrid velocity estimation

The multigrid relaxation method enables us to address three problems :

- Although deterministic the relaxation scheme remains computationally demanding, since it takes several hundred iterations to reach convergence. The use of a multigrid representation allows to propagate the MRF interactions more efficiently between distant points, yielding better performances.

- Only small displacements for which spatial and temporal derivatives interact, can be estimated with the original method. By giving control to low resolutions, large displacements become small in pixels and the interactions between spatial and temporal gradients are restored. This limitation of course does not concern the feature-based measurements, which are not restricted to small displacements.

- The energy functions at different scales do not present the same "landscape". As a matter of fact it has been conjectured, [35] , that in low resolution images the energy landscape is "smoother" and may contain fewer local minima than the energy associated to the original full resolution image. Whereas the deterministic relaxation algorithm gets stuck in the first local minimum near the starting configuration, estimates closer to the global optimum may be expected from a multigrid method starting at a low resolution level where less local minima exist.

The implemented multigrid algorithm consists of a coarse-to-fine strategy, starting from the lowest resolution and propagating the estimates from the coarse scales to the finer ones. A gaussian image pyramid is built up using low-pass filtering and subsampling by a factor of 2 the original images of the sequence. The optical flow at resolution level k is denoted $\vec{\omega}^k$. Three levels of resolution are used in our experiments ($k = 0, 1, 2$). The multigrid algorithm can be described as follows :

- **1** Estimation of the optical flow at the lowest resolution level ($k = 2$) using the original multimodal scheme.

- **2** Repetition and bilinear interpolation of the vectors from the coarse level k to the finer level $k-1$. The interpolation takes into account the location of the motion boundaries. The interpolated field is denoted $\vec{\omega}_0^{k-1}$.

- **3** Estimation of an *incremental* optical flow field $\vec{d\omega}^{k-1}$ at level $k-1$ introducing a modified version of the image flow equation, [16] in the global energy function :

$$\vec{\nabla} f(s + \frac{\vec{\omega}_0^{k-1}}{\Delta t}, t + \Delta t).\vec{d\omega}^{k-1}(s) \; + \; f(s + \frac{\vec{\omega}_0^{k-1}}{\Delta t}, t + \Delta t) - f(s,t) \; = \; 0$$

  The relaxation is performed until convergence at that level (the convergence criterion is the same as in the monoresolution case). The final optical flow field at level $k-1$ is : $\vec{\omega}^{k-1} = \vec{\omega}_0^{k-1} + \vec{d\omega}^{k-1}$. The motion boundaries are estimated using the same energy function as in the monoresolution method.

- **4** If the current level is 0, stop; else $k := k-1$, goto 2.

In the experiments carried out, the same parameters values are used for the potential functions at each level of the image pyramid. It is not clear, how the parameters of the MRF model should vary in a multiresolution structure. Gidas, [23] , recently proposed a coherent mathematical framework for multiscale MRF estimation. Unfortunately the multiresolution energy functions are tractable only for very simple models, such as the Ising model.

The contribution of the multigrid relaxation method is illustrated on a TV sequence called *Mobi* comprising large displacements, several different moving objects

and camera motion. The scene is composed of a rolling ball, a moving toy-train and a calendar undergoing a vertical translation (Fig. 14a). The camera motion corresponds to a panning of the scene, which yields an additional horizontal translation component in the optical flow.

Due to large displacements along with important uniform areas (on the calendar for instance) and sharp edges, the spatial and temporal derivatives interact on a very short range in that case. Therefore the final optical flow field computed by the original monoresolution scheme is not satisfactory (see Fig 14b). The estimates delivered by the multigrid algorithm, with three resolution levels, are presented in Fig. 14c. Visually the optical flow recovered in the multiresolution case is closer to the real underlying motion (see for example the apparent diagonal translation on the calendar). Table 1 shows the number of iterations required at each resolution level to reach convergence. The total iteration number in the multiresolution case corresponds to an equivalent number of 56 iterations at full resolution. This is twice less than for the monoresolution algorithm which requires 111 iterations to converge to a result of lower quality.

	level	Iteration Nb	Equ. Iter.
Multiresolution	0	43	56
	1	37	
	2	59	
Monoresolution	0	111	111

Table 1 : Number of iterations in the mono and multiresolution scheme (Mobi sequence).

Equ. Iter. : computational equivalent of one complete sweep through the image at the finest resolution.

5.2 Multiframe motion analysis

Obviously the optical flow fields $\vec{\omega}_{t\rightarrow t+\Delta t}$ and $\vec{\omega}_{t+\Delta t\rightarrow t+2\Delta t}$ extracted from the successive frames ($f(.,.,t)$, $f(.,.,t+\Delta t)$) and $f(.,.,t+\Delta t)$, $f(.,.,t+2\Delta t)$ are generally similar, except when they are computed on a cut between two different sequences or when there is an abrupt change in the content of the scene. More precisely, it is reasonable to consider that the displacement $dP_{t\rightarrow t+\Delta t}$ of a given point P between time t and $t+\Delta t$ is close to the displacement $dP_{t+\Delta t\rightarrow t+2\Delta t}$ of the displaced point $P+dP_{t\rightarrow t+\Delta t}$ between time $t+\Delta t$ and $t+2\Delta t$.

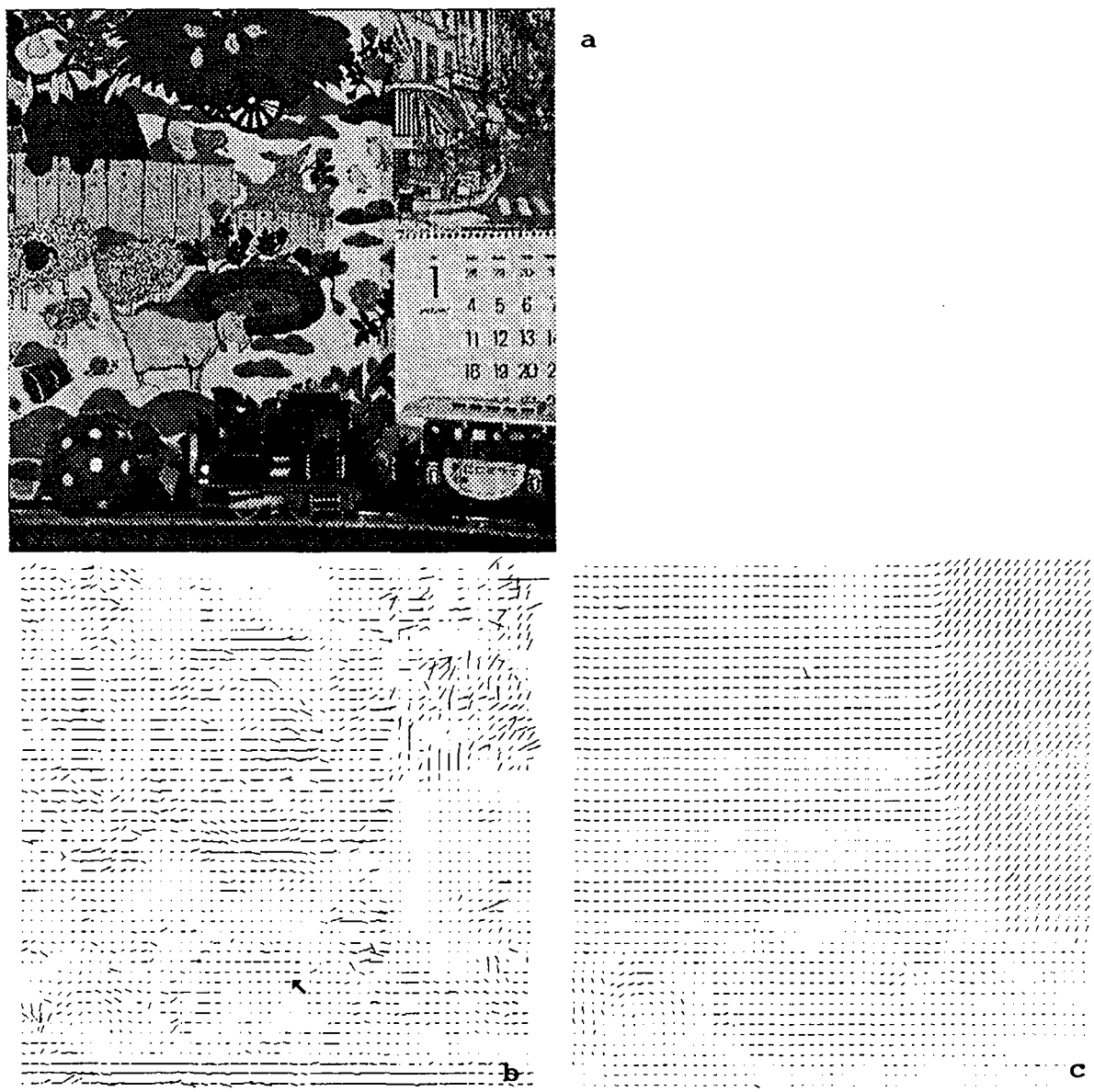


Figure 14 : Multigrid estimation : "Mobi" sequence (by courtesy of CCETT-Rennes).

a) First frame of the original sequence (512×512) : the scene is composed of a rolling ball, a moving toy-train and a calendar undergoing a vertical translation. The camera motion corresponds to a panning of the scene.

b) Optical flow estimation with the original monoresolution scheme (horizontally and vertically subsampled by 10).

c) Optical flow estimation with the multigrid method (horizontally and vertically subsampled by 10).

Therefore, when a long sequence is processed, the number of iterations can be strongly reduced, by using the field obtained on the previous frames to initialize the current optical flow computation. The field used for initialization at time $t + \Delta t$ is obtained from the optical flow field computed between t and $t + \Delta t$ after projection in the direction of motion and interpolation on the discrete image grid.

Initialization	Frame No	Iteration Nb
zero-value field	2	275
	3	324
	4	256
	5	307
	6	387
predicted field	2	.
	3	137
	4	125
	5	104
	6	186

Table 2 : Number of iterations required to reach convergence (zero initialization and initialization by the predicted field) (Interview sequence)

The saving of computation time appears clearly in a comparison with the original scheme (in which an initialization to zero is applied)(see Table 2). The results in Table 2 correspond to the processing of six successive frames extracted from the sequence *Interview*.

Whereas it takes between 250 and 400 iterations to reach convergence in the zero-initialization case, only half of this number is needed in the modified scheme. Other sequences have been processed with similar results.

When there are cuts in the sequence, the modified algorithm does not behave as well of course. One can even expect that in some situations a wrong initialization may lead to a convergence to a local minimum far from the global optimum. To avoid this, efficient methods for detecting cuts or abrupt changes in the scene content are available and could be used in addition (yet, this issue was not investigated here).

6 Concluding Remarks

We have presented a general algorithm for optical flow estimation which is able to jointly handle discontinuities and occlusions in the motion field. It can be interpreted as a generalized regularization approach to the ill-posed problem of optical flow computation. The method has been called multimodal in that it integrates several complementary constraints on the desired solution. Statistical models express the interactions between the different low-level image entities : velocity vectors, motion boundaries, occluding and occluded surfaces, intensity edges and the spatio-temporal variations of the brightness pattern. The motion measurements are based on two complementary constraints : gradient-based and feature-based. The algorithm requires the tuning of only two main parameters which balance the smoothing of the velocity field and the sensitivity of the motion boundary detection.

Experiments have been carried out on a large number of real-world sequences : outdooor and indoor scenes with moving camera, several moving objects and large displacements. One key feature of the described scheme is its ability to handle properly occlusions areas. Indeed in image sequences discontinuities are not only local but exist on large areas corresponding to occluded surfaces. This problem has been addressed here by testing directly the validity of the underlying motion measurement equations. We think this is an efficient way to cope with the general occlusion problem. The experimental results on synthetic sequences clearly demonstrate the advantages of this approach. We think that the multimodal estimation algorithm should be considered as a step toward a comprehensive multiconstraint motion estimator. Such an estimator would enable velocity estimation in very general situations : on textured outdoor scenes as well as on structured man-made environments with long-range or short-range motion.

Yet, a perfect detection of motion boundaries remains difficult : for instance the extracted motion boundaries are sometimes locally broken. This is mainly due to two factors : the quality of intensity edges used as partial support for estimating motion boundaries and the use of a deterministic optimization algorithm which yields suboptimal motion edge configurations.
However, the statistical framework described here seems flexible enough to allow several important extensions. For instance, in the present scheme, only local motion boundaries are determined. An extension toward a region-based motion segmentation algorithm would be of interest in the context of dynamic scene analysis. A partition of a sequence into its constituent moving objects indeed defines a first key-step in many dynamic scene analysis problems.

Another straightforward addition may deal with the multimodal cooperation with other early motion measurements resulting from similarity functions, token tracking or grey-value corners matching, for instance.

Besides, as far as large displacements are concerned, a significant contribution

to multigrid MRF-based relaxation algorithms would be the development of a consistent theoretical framework for multiscale Markov modeling. Up to now, to our knowledge the only satisfactory paradigm, [23] , does not lead to tractable energy functions. As a matter of fact, it is not known how the parameters of the potential functions vary when the scale is changed. A tractable multiscale framework for MRF would undeniably help the statistical multiscale analysis of images (with applications in many other fields such as stereovision, image restoration or segmentation).

The last point is the temporal stability of the extracted motion cues (motion boundaries or regions). In the described algorithm there is no *a priori* modeling of the connections which naturally exist between estimates obtained at different times. It would be of great interest to have also a control on the temporal dimension, in order for example to filter and to track motion cues along the sequence. To this end we currently study temporal extensions to the existing spatial Markov Random Field models. A class of models with temporal neighborhoods has already been introduced in motion detection [7] and motion segmentation, [41] . They appear promising as far as the processing of long sequences is concerned.

Acknowledgments

We would like to thank T. Catudal and E. Nestour for the implementation of the multigrid method and the mutiframe extension of the original algorithm discussed in Section 5.

This work has been partly supported by MRT and CNRS in the context of the PRC Program "Man Machine Interface" under contract PMFE 88F1 548.

References

[1] ADIV G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 7: pages 384–401, July 1985.

[2] AGGARWAL, J.K. and NANDHAKUMAR, N. On the computation of motion from sequences of images - a review. *Proc. IEEE*, Vol. 76, No 8: pages 917–935, 1988.

[3] BARNARD S.T. Stochastic stereo matching over scale. *Int. J. Comp. Vis.*, Vol 3: pages 17–32, 1989.

[4] BESAG, J. On the statistical analysis of dirty pictures. *J. Royal Statist. Soc.*, Vol. 48, Serie B, No 3: pages 259–302, 1986.

[5] BOUTHEMY P. Extracting dense motion information from an image sequence : optic flow estimation and related segmentation issues. In G.E. Taylor, editor, *Kinematic and Dynamic Issues on Sensor Based Control*, pages 223–250, Springer, 1990.

[6] BOUTHEMY, P. A maximum-likelihood framework for determining moving edges. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No 5: pages 499 511, May 1989.

[7] BOUTHEMY P. and LALANDE P. Detection and tracking of moving objects based on a statistical regularization method in space and time. In *Proc. First European Conference on Computer Vision*, pages 307–311, Springer, Antibes, France, April 1990.

[8] BRUSS A.P. and HORN B.K.P. Passive navigation. *Comput. Vision, Graphics, Image Processing*, Vol 21: pages 3–20, 1983.

[9] CHALMOND, B. Image restoration using an estimated Markov model. *Signal Processing*, Vol. 15, No 2: pages 115–129, September 1988.

[10] CHENG J.K. and HUANG T.S. Image registration by matching relational structures. *Pattern Recognition*, Vol 17, No 1: pages 149–159, 1984.

[11] CHOU, P.B. and BROWN, C.M. The theory and practice of bayesian image modeling. *Int. J. Comp. Vis.*, Vol 4: pages 185–210, 1990.

[12] CORNELIUS, N. and KANADE, T. Adapting optical flow to measure object motion in reflectance and x-ray image sequences. In *ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion*, pages 50–58, Toronto, April 1983.

[13] CROWLEY J.L., STELMASZYK P., and DISCOURS C. Measuring image flow by tracking edge-lines. In *Proc. 2nd Int. Conf. Computer Vision*, pages 658 664, Tarpon Springs, Florida, Dec. 1988.

[14] DERICHE, R. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. J. Computer Vision*, pages 167–187, 1987.

[15] DERICHE R. and FAUGERAS O. Tracking line segments. In *Proc. First European Conference on Computer Vision*, pages 259–268, Springer, Antibes, France, April 1990.

[16] ENKELMANN W. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Comput. Vision, Graphics, Image Processing*, Vol 43: pages 150–177, 1988.

[17] FRANCOIS E. and BOUTHEMY P. The derivation of qualitative information in motion analysis. *Image and Vision Computing*, to be published.

[18] GAMBLE E. and POGGIO T. *Visual Integration and Detection of Discontinuities : The Key Role of Intensity Edges.* Technical Report A.I. 970, M.I.T., October 1987.

[19] GEMAN, D., GEMAN, S., GRAFFIGNE C., and PONG D. Boundary detection by constrained optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 12, No 7: pages 609–628, July 1990.

[20] GEMAN, S. and GEMAN, D. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, No 6: pages 721–741, November 1984.

[21] GEMAN, S. and HWANG C.R. Diffusions for global optimization. *SIAM J. Control and Optimization*, Vol. 24, No 5: pages 1031–1043, September 1986.

[22] GENNERT, M.A. and NEGAHDARIPOUR, S. *Relaxing the Brightness Constancy Assumption in Computing Optical Flow.* Technical Report A.I. 975, M.I.T., June 1987.

[23] GIDAS B. A renormalization group approach to image processing problems. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 11 No 2: pages 164-180, February 1989.

[24] HEEGER, D.J. Optical flow using spatiotemporal filters. *Int. J. Comp. Vis.*, Vol 1, No 4: pages 279–302, Jan. 1988.

[25] HEITZ, F. and BOUTHEMY, P. Motion estimation and segmentation : occlusion analysis using markov random field models (in french). In *7th Congrès AFCET-IRIA*, pages 1359–1368, Paris, Nov. 1989.

[26] HEITZ, F. and BOUTHEMY P. Multimodal motion estimation and segmentation using markov random fields. In *10th Int. Conf. Pattern Recognition*, pages 378–383, Atlantic City, June 1990.

[27] HEITZ, F., MAITRE, H., and de COUESSIN, C. Event detection in multisource imaging using contextual estimation. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 1647–1650, Glasgow, May 1989.

[28] HILDRETH, E.C. *The measurement of Visual Motion.* MIT Press, 1984.

[29] HORN, B.K.P. and SCHUNCK, B.G. Determining optical flow. *Artificial Intelligence*, Vol. 17: pages 185–203, 1981.

[30] HSU, Y. Z., NAGEL, H.-H, and REKERS, G. New likelihood test methods for change detection in image sequences. *Comput. Vision, Graphics, Image Processing*, Vol. 26: pages 73–106, 1984.

[31] HUANG T.S., editor. *Image Sequence Processing and Dynamic Scene Analysis.* Volume F2 of *NATO-ASI Series*, Springer, 1983.

[32] HUTCHINSON J., KOCH C., LUO J., and MEAD C. Computing motion using analog and binary resistive networks. *Computer*, Vol. 21: pages 52–63. March 1988.

[33] JENG F.-C. and WOODS J.W. Simulated annealing in compound gaussian random fields. *IEEE Trans. Inf. Theory*, Vol. 36, No 1: pages 94–107, 1990.

[34] KEARNEY, J.K., THOMPSON, W.B., and BOLEY, D.L. Optical flow estimation: an error analysis of gradient-based methods with local optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 9, No 2: pages 229–244, 1987.

[35] KONRAD, J. *Bayesian Estimation of Motion Fields From Image Sequences.* PhD thesis, INRS-Telecommunications, Montreal, Canada, October 1989.

[36] KONRAD, J. and DUBOIS, E. Estimation of image motion fields : bayesian formulation and stochastic solution. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 1072–1075, New-York, 1988.

[37] LITTLE J.J. and GILLETT W.E. Direct evidence for occlusion in stereo and motion. In *Proc. First European Conference on Computer Vision*, pages 336 340, Springer, Antibes, France, April 1990.

[38] MARROQUIN J.L., MITTER S., and POGGIO T. Probabilistic solution of ill-posed problems in computational vision. *J. American Statis. Assoc.*, Vol 82: pages 76–89, 1987.

[39] MILLER M.I., ROYSAM B., SMITH K., and O'SULLIVAN J.A. Representing and computing regular languages on massively parallel networks. *IEEE Trans Neural Networks*, to appear, Jan. 1991.

[40] MITICHE A., WANG Y.F., and AGGARWAL J.K. Experiments in computing optical flow with the gradient-based, multiconstraint method. *Pattern Rec.*, Vol 20, No 2: pages 173–179, 1987.

[41] MURRAY, D.W. and BUXTON, H. Scene segmentation from visual motion using global optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 9, No 2: pages 220–228, March 1987.

[42] NAGEL, H.H. Image sequences - ten (octal) years - from phenomenology towards a theoretical foundation. *Int. J. Pattern Rec. Artif. Intel.*, Vol 2, No 3: pages 459–483, 1988.

[43] NAGEL, H.H. On a constraint equation for the estimation of displacement rates in image sequences. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 11, No 1: pages 13–30, Jan. 1989.

[44] NAGEL, H.H. On the estimation of optical flow : relations between different approaches and some new results. *Artificial Intelligence*, Vol. 33: pages 299–324, 1987.

[45] NAGEL, H.H. and ENKELMANN W. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8: pages 565–593, 1986.

[46] ROUGEE A., LEVY B.C., and WILSKY A.S. Reconstruction of two-dimensional velocity fields as a linear estimation problem. In *Proc. First Int. Conf. Computer Vision*, pages 646–650, London, U.K., June 1987.

[47] SCHUNCK, B.G. The image flow constraint equation. *Computer Vision, Graphics and Image Processing*, Vol. 35: pages 20–46, 1986.

[48] SCHUNCK, B.G. Image flow segmentation and estimation by constraint line clustering. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol 11, No 10: pages 1010–1027, Oct. 1989.

[49] SPOERRI A. and ULLMAN S. The early detection of motion boundaries. In *Proc. First Int. Conf. Computer Vision*, pages 209–218, London, U.K., June 1987.

[50] TERZOPOULOS, D. Image analysis using multigrid relaxation methods. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 8, No 2: pages 129–139, March 1986.

[51] THOMPSON, W.B., MUTCH, K.M., and BERZINS V.A. Dynamic occlusion analysis in optical flow fields. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 7, No 4: pages 374–383, July 1985.

[52] THOMPSON W.B. and PONG T.C. Detecting moving objects. In *Proc. 2nd Int. Conf. Computer Vision*, pages 201–208, London, U.K., June 1987.

[53] VERRI, A. and POGGIO T. Motion field and optical flow: qualitative properties. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No 5: pages 490–498, May 1989.

[54] WAXMAN A.M. and ULLMAN S. Surface structure and three-dimensional motion from image flow kinematics. *Int. J. Robot Res.*, Vol. 4, No 3: pages 72–94, 1985.

[55] YUILLE A.L. and BULTHOFF H. Stereo integration, mean field theory and psychophysics. In *Proc. First European Conference on Computer Vision*, pages 73–82, Springer, Antibes, France, April 1990.

[56] YUILLE A.L. and GRZYWACZ N.M. The motion coherence theory. In *Proc. 2nd Int. Conf. Computer Vision*, pages 344–353, Tarpon Springs, Florida, Dec. 1988.

LISTE DES DERNIERES PUBLICATIONS INTERNES IRISA

PI 550 **MULTISCALE SIGNAL PROCESSING : FROM QMF TO WAVELETS**
Albert BENVENISTE
Septembre 1990, 28 Pages.

PI 551 **ON THE TRANSITION GRAPHS OF AUTOMATA AND GRAMMARS**
Didier CAUCAL, Roland MONFORT
Septembre 1990, 46 Pages.

PI 552 **ERREURS DE CALCUL DES ORDINATEURS**
Jocelyne ERHEL
Septembre 1990, 58 Pages.

PI 553 **SEQUENTIAL FUNCTIONS**
Boubakar GAMATIE, Octobre 1990, 16 Pages.

PI 554 **ANALYSE DE LA FORME D'UN COEFFICIENT D'ASSOCIATION ENTRE VARIABLES QUALITATIVES**
Mohamed OUALI ALLAH
Octobre 1990, 26 Pages.

PI 555 **APPROXIMATION BY NONLINEAR WAVELET NETWORKS**
Qinghua ZHANG, Albert BENVENISTE
Octobre 1990, 16 Pages.

PI 556 **CONCEPTION ET INTEGRATION D'UN CORRELATEUR SYSTOLIQUE**
Catherine DEZAN, Eric GAUTRIN, Patrice QUINTON
Novembre 1990, 16 Pages.

PI 557 **VARIATIONAL APPROACH OF A MAGNETIC SHAPING PROBLEM**
Michel CROUZEIX
Novembre 1990, 14 Pages.

PI 558 **THE DAVIDSON METHOD**
Michel CROUZEIX, Bernard PHILIPPE et Miloud SADKANE
Novembre 1990, 22 Pages.

PI 559 **A DISTRIBUTED SOLUTION TO THE ℓ-OUT OF-M RESOURCES ALLOCATION PROBLEM**
Michel RAYNAL
Novembre 1990, 18 Pages.

PI 560 **A SIMPLE TAXONOMY FOR DISTRIBUTED MUTUAL EXCLUSION ALGORITHMS**
Michel RAYNAL
Novembre 1990, 8 Pages.

PI 561 **MULTIMODAL ESTIMATION OF DISCONTINUOUS OPTICAL FLOW USING MARKOV RANDOM FIELDS**
Fabrice HEITZ, Patrick BOUTHEMY
Novembre 1990, 50 Pages.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique.

ISSN 0249-6399